

## Bayesian mode regression

Journal:	<i>Scandinavian Journal of Statistics</i>
Manuscript ID:	SJS-14-188
Manuscript Type:	Paper
Date Submitted by the Author:	14-Aug-2014
Complete List of Authors:	Yu, Keming; Brunel University, Mathematics

SCHOLARONE™  
Manuscripts

Review

# Bayesian Mode Regression

Keming Yu, *Brunel University, London, UB8 3PH, UK*

Katerina Aristodemou, *Brunel University, London, UB8 3PH, UK*

Zudi Lu, *University of Southampton, Southampton, SO17 1BJ, UK*

## Abstract

Like mean, quantile and variance, mode is also an important measure of central tendency of a distribution. Many practical questions, particularly in the analysis of big data, such as “Which element (gene or file or signal) is the most typical one among all elements in a network?” are directly related to mode. Mode regression, which provides a convenient summary of how the regressors affect the conditional mode, is totally different from other models based on conditional mean or conditional quantile or conditional variance. Some inference methods for mode regression exist but none of them is from the Bayesian perspective. This paper introduces Bayesian mode regression by exploring three different approaches, including their theoretic properties. The proposed approaches are illustrated using simulated datasets and a real data set.

**Keywords:** Bayesian inference; Empirical likelihood; Mode regression

## 1. INTRODUCTION

Mode, the most likely value of a distribution, has wide applications in biology, astronomy, economics and finance. In these fields, it is not uncommon to encounter data distributions that are skewed or contain outliers. In those cases, the arithmetic mean may not be an appropriate statistic to represent the center of location of the data. Alternative statistics with less bias are the median and the mode. The mean or the median of two densities may be identical, while the shapes of the two densities can be quite different. The mode preserves some of the important features, such as wiggles, of the underlying distribution function, whereas the mean and the median tend to average out the data.

1  
2  
3 The mode has been used in modern science to identify the most frequent or the most typical  
4 element in certain network systems (Hedges and Shah (2003), Heckman et al. (2001), Kumar and  
5 Hedges (1998), Markov et al. (1997)). Mode estimation has attracted significant attention in the  
6 statistics literature for decades by various authors [Yasukawa (1926), Parzen (1962), Grenander  
7 (1965), Eddy (1980), Bickel and Fan (1996), Birgé (1997), Berlinet et al. (1998) and Meyer (2001)  
8 among others]. Moreover, identifying the typical value or pattern could be one of the most efficient  
9 statistical approaches for the analysis of big data.  
10  
11

12  
13 However, mode estimation is more difficult than estimating the mean or the median. The  
14 mode estimator is often defined as the maximum of the estimated distribution density, typically  
15 under nonparametric kernel estimation. Conditional mode estimation is typically carried out by  
16 conditional density estimation via different nonparametric methods [see for example Gasser et al.  
17 (1998), Hall and Huang (2001) and Hall et al. (2001), Brunner (1992), Ho (2006), Dunson et al.  
18 (2007)].  
19  
20

21  
22 However, these nonparametric conditional density-based mode regression models do not provide  
23 a direct estimate of the conditional mode. The problem with these methods is twofold: the estima-  
24 tion of the conditional density may suffer from the well-known “curse of dimensionality” and, it is  
25 hard to describe and interpret the estimated conditional mode in terms of predictors or covariates.  
26  
27

28  
29 Direct inference for mode regression was explored by Lee first in 1989, Lee (1989), and then  
30 in 1993, Lee (1993). However, it has not been well-applied due to lack of proper inference tools.  
31 Recently, Kemp and Santos Silva (2012) relaxed Lee’s restriction on truncated dependent variables  
32 and employed alternative kernel estimation. However, their regression coefficient estimator has slow  
33 convergence rate, involves bandwidth selection and provides only approximate Normal confidence  
34 intervals. Furthermore, Yao and Li (2013) proposed an Expectation-Maximisation algorithm in  
35 order to estimate the regression coefficients of the modal linear regression. These methods involve  
36 either semiparametric or nonparametric estimation methods. A direct Bayesian method for mode  
37 regression is not available even though there is a clear practical motivation from this perspective.  
38  
39

40  
41 In conventional regression models, the method of least squares is usually applied to investigate  
42 the effect of the predictor variables on the conditional mean of the response variable. However, in  
43 the presence of outliers, the mean is pulled in the direction of the tail, making mean regression a less  
44 representative method of analysis. Mode regression, on the other hand, is robust to the presence  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 of outliers. Quantile regression is an alternative approach to estimate models with skewed data, as  
4  
5 it can provide a complete picture of the conditional distribution of the response variable given the  
6  
7 covariates. However, it cannot reveal any information about the typical value (mode).  
8

9 Take the analysis of the adult Body Mass Index (BMI) used in this paper as an example. BMI,  
10 defined by  $BMI = \frac{weight(kg)}{height^2(cm)}$ , is a measure of the relative weight and is used in a wide variety of  
11  
12 contexts as a simple method to assess how much an individual's body weight deviates from what  
13  
14 is normal or desirable for a person of his or her height. Such analysis is important as it is well-  
15  
16 known that obesity has overtaken smoking as the biggest threat to people's health, in particular  
17  
18 for middle-aged and old adults.  
19

20 The dataset used in this paper to demonstrate mode regression is taken from the Health Survey  
21  
22 for England (HSE) 2011 teaching dataset. The Health Survey for England is a series of annual  
23  
24 surveys about the health of people living in England, commissioned by the Department of Health.  
25  
26 The sample contains observations for 4,138 individuals (1,814 males and 2,324 females) with two  
27  
28 thirds being older than 40 years old. A BMI of  $27kg/m^2$  for middle-aged and old adults can be  
29  
30 classified as the cut-off point of unhealthy weight. An interesting question is how some covariates,  
31  
32 such as units of alcohol and portions of fruit/vegetables consumed keep one's BMI in the healthy  
33  
34 range. It would be safe to assume that the BMI for the majority of people in the data example falls  
35  
36 in the desirable BMI range. Indeed, the typical BMI for the whole sample as well as separately  
37  
38 for men or women are below  $27kg/m^2$ , but the corresponding mean BMI and median BMI were  
39  
40 near or greater than  $27kg/m^2$ . Therefore, employing mode regression is preferable than mean and  
41  
42 quantile regression for answering this scientific question.

43 In this paper we introduce a fully Bayesian framework for direct mode regression inference  
44  
45 by using three approaches: a parametric Bayesian method, a nonparametric Bayesian method  
46  
47 and an empirical likelihood based Bayesian method. The remainder of the paper is organized as  
48  
49 follows. Section 2 introduces the three approaches, describes the theoretical and computational  
50  
51 framework of these methods and gives their mathematical justification. In Section 3 we illustrate  
52  
53 the proposed methods through two simulated case-studies and a real example. We conclude with  
54  
55 a short discussion in Section 4.  
56  
57  
58  
59  
60

## 2. BAYESIAN MODE REGRESSION

### 2.1. Mode Estimation

Consider an arbitrary random variable  $Z$ , with distribution function  $F_Z(z)$  and density function  $f_Z(z)$ . Let  $K(Z; \cdot)$  be the *step-loss function* (Manski (1991)) such as,

$$K(Z; \mu) = I \left[ \frac{|Z - \mu|}{\sigma} > 1 \right], \quad (2.1)$$

with  $\sigma > 0$  and  $I[A]$  being the indicator function of event  $A$ . If  $f_Z(z)$  is symmetric around  $\mu$  or if  $\mu$  is the middle value of the interval of length  $2\sigma$  that captures the most probability under  $F_Z(z)$ , then

$$\hat{\mu} = \operatorname{argmin}_{\mu} E\{K(Z; \mu)\}$$

is the mode of  $Z$ .

Therefore, given a sample  $\{Z_1, \dots, Z_n\}$  from  $Z$ , let  $\hat{\mu}$  be the estimator of the mode of  $Z$ , then,

$$\hat{\mu} = \operatorname{argmin}_{\mu} \sum_{i=1}^n I[|Z_i - \mu| > \sigma].$$

Generally, if we define a uniform density function  $f_{\sigma}(u)$  over the interval  $(\mu - \sigma, \mu + \sigma)$  as

$$f_{\sigma}(u; \mu) = \frac{1}{2\sigma} I(|u - \mu| \leq \sigma), \quad (2.2)$$

where  $-\infty < \mu < \infty$  and  $\sigma > 0$  are the location parameter and scale parameter respectively, then clearly the mode of  $Z$  can be estimated by

$$\hat{\mu} = \operatorname{argmax}_{\mu} \prod_{i=1}^n f_{\sigma}(Z_i; \mu). \quad (2.3)$$

### 2.2. Parametric Bayesian Mode Regression

Let  $\operatorname{mode}(y|\mathbf{x}) = \mathbf{x}'\boldsymbol{\beta}$  be the conditional mode of  $Y$  given  $X = x$ . A standard regression model to formulate the mode regression could be as:

$$y = \mathbf{x}'\boldsymbol{\beta} + \epsilon, \quad (2.4)$$

with  $\text{mode}(\epsilon|\mathbf{x}) = 0$  of model error  $\epsilon$ .

Lee (1989,1993) showed that, given a sample  $\{(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_n, y_n)\}$  from  $(\mathbf{x}, y)$ , the classical mode regression estimator,  $\hat{\boldsymbol{\beta}}$  is given by:

$$\hat{\boldsymbol{\beta}} = \operatorname{argmax}_{\boldsymbol{\beta}} \frac{1}{n} \sum_{i=1}^n I[|y_i - \mathbf{x}'_i \boldsymbol{\beta}| \leq \sigma]. \quad (2.5)$$

Therefore, using equation (2.3),  $\hat{\boldsymbol{\beta}}$  can be regarded as the maximum likelihood estimator of the “working” likelihood function

$$L(y|\boldsymbol{\beta}, \sigma) = \prod_{i=1}^n f_{\sigma}(y_i; \mathbf{x}'_i \boldsymbol{\beta}). \quad (2.6)$$

Now a natural joint posterior distribution of the unknown model parameters,  $\boldsymbol{\beta}$  and  $\sigma$  under a Bayesian framework is given by

$$\pi(\boldsymbol{\beta}, \sigma|y) \propto L(y|\boldsymbol{\beta}, \sigma) \pi(\boldsymbol{\beta}, \sigma), \quad (2.7)$$

where  $\pi(\boldsymbol{\beta}, \sigma)$  is the joint prior distribution of  $\boldsymbol{\beta}$  and  $\sigma$ .

The Bayesian mode regression estimates, denoted as  $\hat{\boldsymbol{\beta}}_B$  can be obtained using the marginal posterior distribution of  $\boldsymbol{\beta}$ , given by

$$\pi(\boldsymbol{\beta}|y) = \int \pi(\boldsymbol{\beta}, \sigma|y) d\sigma, \quad (2.8)$$

In a similar manner, an estimate of  $\sigma$ , denoted as  $\hat{\sigma}$ , can be obtained using the marginal posterior distribution of  $\sigma$ ,

$$\pi(\sigma|y) = \int \pi(\boldsymbol{\beta}, \sigma|y) d\boldsymbol{\beta}, \quad (2.9)$$

Although a standard conjugate prior distribution is not available for the mode regression formulation, Markov Chain Monte Carlo (MCMC) methods may be used for extracting the posterior distributions of both  $\boldsymbol{\beta}$  and  $\sigma$ .

### 2.3. Estimation of Covariance Matrix of Classical Estimates

Under the classical approaches of Lee (1989, 1993) and Kemp and Santos Silva (2012), the covariance matrix,  $cov\{\hat{\beta}\}$  of the classical estimator  $\hat{\beta}$  and its inverse are often required but difficult to estimate or compute numerically, especially under small or moderate samples. A by-product of the proposed Bayesian approach is that using the MCMC posterior sample leads to a natural and efficient estimation of  $cov\{\hat{\beta}\}$  and other asymptotic quantities of  $\hat{\beta}$ .

In fact, a MCMC scheme constructs a Markov chain whose equilibrium distribution is the joint posterior,  $p(\beta|data)$ . After running the Markov chain for a burn-in period, one obtains samples from the limiting distribution, provided that the Markov chain has converged. Given that the chain has converged, the frequency of appearance of the parameters in the Markov chain represents their posterior distribution. An informative full density distribution of the model parameters is readily obtained rather than a single point estimate as in the classical approach.

When a Markov chain,  $S$ , is drawn from the posterior distribution,  $p(\beta|data)$ :  $S = (\beta^{(1)}, \beta^{(2)}, \dots, \beta^{(N)})$ , where  $N$  is the number of draws after burn-in, a consistent estimate of the inverse of the covariance matrix  $cov\{\hat{\beta}\}$  can be obtained by multiplying by  $N$  the variance-covariance matrix of this MCMC sequence (Chernozhukov and Hong 2003).

### 2.4. Prior Section and Proper posteriors

In this section first we demonstrate that almost all priors for  $(\beta, \sigma)$  could be used and yield a proper joint posterior. In fact we have the following theorem.

**Thm 2.1.** *Given the mode regression (2.4) and the ‘working’ likelihood (2.6), if the joint prior distribution  $\pi(\beta, \sigma)$  follows one of the following three choices:*

- (1)  $\pi(\beta, \sigma) \propto 1$  (totally non-informative prior)
- (2)  $\pi(\beta, \sigma) = \pi(\beta) \pi(\sigma|\beta)$  and one of  $\pi(\beta)$  and  $\pi(\sigma|\beta) \propto 1$  and the other is a proper prior,
- (3)  $\pi(\beta, \sigma) = \pi(\beta) \pi(\sigma|\beta)$  and both  $\pi(\beta)$  and  $\pi(\sigma|\beta)$  are proper priors,

then the posterior distribution of  $\beta$  and  $\sigma$ ,  $\pi(\beta, \sigma|\mathbf{y})$ , will be a proper distribution. In other words

$$0 < \int \pi(\beta, \sigma|\mathbf{y}) d\beta d\sigma < \infty,$$

or, equivalently,

$$0 < \int L(\mathbf{y}|\boldsymbol{\beta}, \sigma) \pi(\boldsymbol{\beta}, \sigma) d\boldsymbol{\beta} d\sigma < \infty.$$

The proof can be found in the Appendix.

In practice one usually assumes that the components of  $\boldsymbol{\beta}$  have independent prior distributions which is a special case of the above theorem.

## 2.5. One practical selection of prior on $\sigma$

If the conditional distribution is strictly unimodal and symmetric or if the regressors affect only the location of the distribution, then a consistent estimate of the mode can be obtained with a fixed  $\sigma$  (Lee (1989)). In practice, however, data with such characteristics is relatively rare. In addition, in such cases the added value of mode regression is rather limited as the mode coincides with the mean and the median. To extend mode regression to more interesting applications  $\sigma$  must be allowed to approach zero as the sample size goes to infinity.

A suitable prior distribution for  $\sigma$  would be one with a positive support. To this end it is proposed to use either a Uniform( $w_1, w_2$ ) or a Gamma distribution with mean  $w_i$ , where, in both cases  $w_i$  can be determined using one of the following options, commonly used in bandwidth selection methods for kernel density estimation:

- The empirical rule, which states that, given a symmetric distribution, approximately 99.7% of the data values fall within three standard deviations (sd) of the mean, therefore,  $w_i = 3sd$ ;
- Variations of Silverman's plug-in estimate for the bandwidth (Silverman (1986)), in which  $w_i = 1.3643\delta n^{-0.2}[\min(\widehat{sd}, IQR/1.349)]$ , where,  $IQR$  is the sample inter quantile range and  $\delta = 1.3510$  for a uniform kernel. To cover data with large number of outliers  $IQR/1.349$  can be replaced by  $1.4826MAD$ , where  $MAD$  is the median absolute deviation.

Alternatively, as the next section demonstrates, a more flexible model can be developed by relaxing the distributional assumption on the prior for  $\sigma$  using a Dirichlet process prior. This leads to a flexible nonparametric mixture model. The method is nonparametric in the sense that it is not assumed that the prior belongs to any fixed class of distributions.



## 2.6. Nonparametric Bayesian method

In this section, we formulate a nonparametric Bayesian mode regression model to avoid critical dependence on the mode uniform distribution assumption thus to address the issue of misspecification that may arise under the parametric Bayesian method.

A density  $f(\cdot)$  on  $\mathbb{R}^+$  is non-increasing if and only if there exists a distribution function  $G$  such that  $f(x|G) = \int \sigma^{-1} I_{[0 < x < \sigma]} dG(\sigma)$  (Feller 1971). Therefore, any unknown density  $f(\cdot)$  (with mode  $\theta$ ), symmetric or not, can be represented as a scale mixture of symmetric uniform distributions, that is

$$f(x|\theta, G) = \int \frac{1}{2\sigma} I_{[-\sigma < x - \theta < \sigma]} dG(\sigma), \quad (2.10)$$

where  $G$  is the mixing distribution supported on  $\mathbb{R}^+$ .

Then, a nonparametric Bayesian mode regression model can be expressed in the hierarchical form

$$\begin{aligned} y_i | \beta, \sigma_i &\stackrel{ind}{\sim} f(y_i - x'_i \beta; \sigma_i), i = 1 \cdots n \\ \sigma_i | G &\stackrel{iid}{\sim} G, i = 1 \cdots n \\ G | M, d &\sim DP(M, G_0(\cdot, d)) \\ \beta, M, d &\sim p(\beta), p(M), p(d), \end{aligned} \quad (2.11)$$

where,  $G$  is the mixing distribution, with base distribution  $G_0$  and concentration parameter  $M$  and  $f(y_i - x'_i \beta; \sigma_i) = \frac{1}{2\sigma} I_{[-\sigma < y_i - x'_i \beta < \sigma]}$  is the density of a uniform distribution on  $(-\sigma, \sigma)$ .

We take a uniform distribution as the base distribution,  $G_0$ , uniform prior for  $M$  and we choose non-informative Normal priors for all the components of  $\beta$ .

## 2.7. Empirical Likelihood based Bayesian Method

In addition to parametric and nonparametric likelihood, an empirical likelihood based method could be an alternative for Bayesian mode regression. To derive an empirical likelihood for mode regression we begin with notations and a moment restriction. Lee (1993) generalized the mode regression estimator of Lee (1989),  $\hat{\beta} = \operatorname{argmin}_{\beta} E\{L(Y - x' \beta)\}$ , by using the triangular kernel  $L(Y; \mu) = \{(\sigma^2 - (Y - \mu)^2) I[|Y - \mu| < \sigma]\}$ .

Therefore, the moment restriction for the empirical likelihood can be obtained by the derivative

1  
2  
3  $\frac{\partial}{\partial \mu} L(Y; \mu) = 2(Y - \mu)I[|Y - \mu| < \sigma]$ . Let  $l(Y; \mu)$  be the ‘derivative’ of  $L(.; \mu)$  with respect  
4 to  $\mu$ , then the mode,  $\mu$ , of  $Y$  satisfies the moment restriction  $E(l(Y; \mu)) = 0$ , where  $l(Y; \mu) =$   
5  $(Y - \mu)I(|Y - \mu| < \sigma)$ .  
6  
7

8  
9 Under an empirical likelihood for mode regression  $\mu = \mathbf{x}'\boldsymbol{\beta}$ , thus for any proposed  $\boldsymbol{\beta}$  to estimate  
10 the true  $p$  dimensional  $\boldsymbol{\beta}_0$  via empirical likelihood, we use the vector estimating functions  $g(X, Y, \boldsymbol{\beta})$   
11 with component  $g_j(X, Y, \boldsymbol{\beta}) = l(Y; \boldsymbol{\beta}'X) X_j$  for  $j = 1, \dots, p$ . Then, the profile empirical likelihood  
12 ratio is given by  
13  
14  
15

$$16 \quad \mathfrak{R}(\boldsymbol{\beta}) = \max\left\{\prod_{i=1}^n (n p_i) \mid \sum_{i=1}^n p_i g(X_i, Y_i, \boldsymbol{\beta}) = 0, p_i \geq 0, \sum_{i=1}^n p_i = 1\right\}.$$

17  
18  
19  
20  
21 By a standard Lagrange multiplier argument we have  
22

$$23 \quad \mathfrak{R}(\boldsymbol{\beta}) = \prod_{i=1}^n \{n p_i(\boldsymbol{\beta})\}, \quad (2.12)$$

24  
25 with the weights  $p_i(\boldsymbol{\beta}) = \frac{1}{n(1 + \hat{\lambda}(\boldsymbol{\beta})'g(X_i, Y_i, \boldsymbol{\beta}))}$ , where the Lagrange multiplier  $\hat{\lambda}(\boldsymbol{\beta})$  is the solution of  
26  
27  
28  $\lambda$  to the following equation  
29

$$30 \quad \sum_{i=1}^n \frac{g(X_i, Y_i, \boldsymbol{\beta})}{1 + \lambda^T g(X_i, Y_i, \boldsymbol{\beta})} = 0. \quad (2.13)$$

31  
32 According to [Qin and Lawless \(1994\)](#), among others, the existence and uniqueness of  $\hat{\lambda}(\boldsymbol{\beta})$  are  
33 guaranteed when the following two conditions are satisfied: (1) zero belongs the convex hull of  
34  $\{g(X_i, Y_i, \boldsymbol{\beta}), i = 1, \dots, n\}$  and (2) the matrix  $\sum_{i=1}^n \{g(X_i, Y_i, \boldsymbol{\beta})g(X_i, Y_i, \boldsymbol{\beta})'\}$  is positive definite.  
35

36 Under Bayesian inference we consider the empirical likelihood function  $\mathfrak{R}(\boldsymbol{\beta})/n^n = \prod_{i=1}^n \{p_i(\boldsymbol{\beta})\}$ ,  
37 which can be combined with a prior specification  $\pi(\boldsymbol{\beta})$  on the parameter  $\boldsymbol{\beta}$  to obtain the posterior  
38 distribution  
39  
40

$$41 \quad \pi(\boldsymbol{\beta} | data) \propto \pi(\boldsymbol{\beta}) \mathfrak{R}(\boldsymbol{\beta}).$$

## 42 2.8. Asymptotic Properties of Bayesian Empirical Likelihood

43  
44 Before establishing the asymptotic normality of the empirical likelihood-based Bayesian mode  
45 regression parameter estimates, the consistency of the empirical likelihood estimator must be es-  
46 tablished, which is a necessary condition for the asymptotic normality of the posterior. Since the  
47 criterion function  $g(X, Y, \boldsymbol{\beta})$  results in a non-smooth estimating equations, a similar method to  
48  
49

the one used by Molanes Lopez et al. (2009), among others, is employed to derive the asymptotic results.

Let  $\hat{\beta} = \operatorname{argmax}_{\beta} \mathfrak{R}(\beta)$  be the maximum empirical likelihood estimator (MELE) in a compact set of parameter space which contains the true parameter  $\beta_0$ . Then note that the criterion function  $g(X, Y, \beta)$  can be regarded as a special case of M-estimators as discussed in Chapter 5 of Van der Vaart (1998) and satisfies the conditions of theorem 5.7 in the book. Under some regular conditions imposed on the marginal distribution of  $X$  and on the conditional distribution of  $Y$  given  $X$ , such as uniformly continuous and bounded, and since both  $E\{g(X, Y, \beta)\}$  and  $E\{g(X, Y, \beta)g(X, Y, \beta)'\} > 0$  are sufficiently smooth in a compact set of parameter space, which contains  $\beta_0$ , the consistency condition  $C_3$  of Molanes Lopez et al. (2009) holds. Then the consistency of empirical likelihood estimates is established. Specifically, a rigorous statement of the conditions and theorem is as follows:

**Assumption 1.** There exists a neighborhood  $\mathcal{N}$  of  $\beta_0$  such that  $P(\mathfrak{R}(\beta) > 0) \rightarrow 1$  for any  $\beta \in \mathcal{N}$ , as  $n \rightarrow \infty$ .

**Assumption 2.** The distribution function  $G_X$  of  $X$  has bounded support  $\mathcal{X}$ .

**Assumption 3.** The conditional distribution  $F_X(t)$  of  $Y$  given  $X$  is twice continuously differentiable in  $t$  for all  $X \in \mathcal{X}$ .

**Assumption 4.** At any  $X \in \mathcal{X}$ , the conditional density function  $F'_X(t) = f_X(t) > 0$  for  $t$  in a neighborhood of  $\beta'_0 X$ .

**Assumption 5.**  $E\{g(X, Y, \beta_0)g(X, Y, \beta_0)'\} > 0$  is positive definite.

**Thm 2.2.** Under Assumptions 1–5, the MELE  $\hat{\beta}$  is a consistent estimator of  $\beta_0$ .

Assumptions 1-5 are standard conditions in this kind of asymptotic problems. For example, these conditions are basically similar to Assumptions 3.1-3.5 of Yang and He (2012, pp. 1110) for Bayesian empirical likelihood quantile regression. Assumption 1 is to guarantee that the interior of the convex hull of  $\{g(X_i, Y_i, \beta) : i = 1, \dots, n\}$  for  $\beta \in \mathcal{N}$  contains the vector of zeros with probability tending to one. Assumption 4 ensures that  $\beta_0$  is indeed the unique solution for  $Eg(X, Y, \beta) = 0$ . The proof of Theorem 2.2 is sketched in the Appendix.

The asymptotic normality of the posterior distribution  $\pi(\beta|data)$  could be established using the fact that the empirical log-likelihood ratio for  $\beta$  is well approximated by certain quadratics in the

sense of Lemma 6 of [Molanes Lopez et al. \(2009\)](#) so that,

$$\Gamma_n(\boldsymbol{\beta}) \equiv -n^{-1} \sum_{i=1}^n \log(1 + \hat{\lambda}(\boldsymbol{\beta})' g(X_i, Y_i, \boldsymbol{\beta})) \quad (2.14)$$

$$\begin{aligned} &= -\frac{1}{2}(\boldsymbol{\beta} - \boldsymbol{\beta}_0)' V_{12}' V_{11}^{-1} V_{12}(\boldsymbol{\beta} - \boldsymbol{\beta}_0) + n^{-1/2}(\boldsymbol{\beta} - \boldsymbol{\beta}_0)' V_{12}' V_{11}^{-1} W_n \\ &\quad - \frac{1}{2} n^{-1} W_n' V_{11}^{-1} W_n + o_P(n^{-1}), \end{aligned} \quad (2.15)$$

with matrices  $V_{11} = (E\{g_j(X, Y, \boldsymbol{\beta}_0) g_k(X, Y, \boldsymbol{\beta}_0)'\})_{j,k=1}^p$ ,  $V_{12} = (-\frac{\partial}{\partial \beta_k} E\{g_j(X, Y, \boldsymbol{\beta})\} |_{\boldsymbol{\beta}=\boldsymbol{\beta}_0})_{j,k=1}^p$ , and vector  $W_n = n^{-1/2} \sum_{i=1}^n g(X_i, Y_i, \boldsymbol{\beta}_0)$ .

Specifically, we make one more assumption on the prior specification  $\pi(\boldsymbol{\beta})$ .

**Assumption 6.**  $\log\{\pi(\boldsymbol{\beta})\}$  has bounded first derivative in a neighborhood of  $\boldsymbol{\beta}_0$ .

Then from  $\log \mathfrak{R}(\boldsymbol{\beta}) = n\Gamma_n(\boldsymbol{\beta})$  we have

**Thm 2.3.** *Under Assumptions 1-6, the posterior density of  $\boldsymbol{\beta}$  has the following expansion on any sequence of sets  $\{\boldsymbol{\beta} : \boldsymbol{\beta} - \boldsymbol{\beta}_0 = O(n^{-1/2})\}$ ,*

$$\pi(\boldsymbol{\beta} | \text{data}) = \pi(\boldsymbol{\beta}) \mathfrak{R}(\boldsymbol{\beta}) \propto \exp\left\{-\frac{1}{2}(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})' I_n(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}) + Q_n\right\} \quad (2.16)$$

with  $I_n = nV_{12}' V_{11}^{-1} V_{12}$  and empirical likelihood estimate  $\hat{\boldsymbol{\beta}}$  and  $Q_n = o_p(1)$ . When  $I_n$  is positive definite, we have  $I_n^{1/2}(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})$  converging in distribution to  $N(0, I)$ .

The proof of Theorem 2.3 is sketched in the Appendix.

We finally remark that, as similarly remarked for quantile regression by Remark 3.2 of [Yang and He \(2012, pp. 1110\)](#), the posterior will be improper for flat priors on  $\boldsymbol{\beta}$  in the Bayesian empirical likelihood approach for our mode regression, and therefore we should avoid using flat priors on  $\boldsymbol{\beta}$ .

In the case of the prior distribution shrinking with  $n$ , we may use  $\pi_n(\boldsymbol{\beta})$  satisfying condition similar to Assumption 3.7 of [Yang and He \(2012\)](#) as priors for our mode regression; see Theorem 3.3 of [Yang and He \(2012\)](#) for details.

### 3. NUMERICAL EXPERIMENTS

In this section we demonstrate our approach to Bayesian mode regression through two simulated and one real examples. For the real example we consider a dataset which investigates how factors

such as gender, age, consumption of alcohol, consumption of fruit and vegetables and smoking can affect the body mass index (BMI).

### 3.1. Simulation Example 1

We consider a simulated data from the model

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i, \quad (3.1)$$

where  $x_i \sim N(0, 1)$ ,  $i = 1, \dots, n$  for  $n = 50, 100, 200$  and  $\beta = (1, 2)$ . The following three specifications were considered for the model error  $\epsilon$ :

- Case 1: the standard Normal distribution,  $\epsilon_i \sim N(0, 1)$  - a symmetric error distribution.
- Case 2: a Fisher's Z distribution,  $\epsilon_i \sim 1/2 \log Z$  with  $Z \sim F_{2,2}$  - a skewed error distribution.
- Case 3: a Normal distribution with normally distributed outliers (contaminants) centred at twice the distance between the true mode and the 99<sup>th</sup> percentile of the original Normal distribution and accounting for 20% of the total data points,  $\epsilon_i \sim 0.80N(0, \frac{1}{4}) + 0.20N(2.5, \frac{1}{4})$  (Hedges and Shah (2003)) - an asymmetric error distribution.

We fit parametric Bayesian mode regression (labeled PBMR) for all the cases above. Then for demonstration and comparison purposes we fit empirical likelihood based Bayesian mode regression (labeled ELBMR) for case 2 and nonparametric Bayesian mode regression (labeled NBMR) for case 3.

For the PBMR and ELBMR models, independent Normal distributions were used as priors of each component of  $\beta$ , where the mean and standard derivation of the Normal prior are given by the classical estimator of Lee (1989, 1993) and its estimated standard error respectively. Realisations were simulated from the posterior distributions by means of a single-component Metropolis-Hastings algorithm. Each of the parameters was updated using a random-walk Metropolis algorithm with a Gaussian proposal density centred at the current state of the chain. The variance of the proposal density was chosen to provide an acceptance rate close to the optimal acceptance rate as defined in Roberts and Rosenthal (2001). Convergence was assessed using time series plots and the R package

Table 1: Simulation Example 1: True parameter values (T.V.) and their posterior means, standard deviations (S.D.) and 95% credible intervals (C.I)

n		PBMR				ELBMR				NBMR	
		Normal		Skewed		Asymmetric		Skewed		Asymmetric	
		$\beta_0$	$\beta_1$	$\beta_0$	$\beta_1$	$\beta_0$	$\beta_1$	$\beta_0$	$\beta_1$	$\beta_0$	$\beta_1$
50	T.V	1	2	1	2	1	2	1	2	1	2
	Mean	0.92	2.00	1.07	2.01	0.96	2.02	1.01	2.00	1.09	1.94
	S.D.	0.78	0.77	0.78	0.49	0.34	0.24	0.01	0.01	0.24	0.19
	95%CI	(-0.6,2.1)	(0.5,3.3)	(-0.3,2.6)	(1.2,3.1)	(0.4,1.7)	(1.6,2.5)	(0.99,1.02)	(1.99,2.01)	(0.7,1.5)	(1.5,2.3)
100	T.V	1	2	1	2	1	2	1	2	1	2
	Mean	1.01	2.10	0.95	1.89	1.06	1.94	1.01	2.00	1.06	2.00
	S.D.	0.18	0.25	0.52	0.37	0.98	0.76	0.01	2	0.14	0.12
	95%CI	(0.6,1.3)	(1.6,2.6)	(0.0,1.9)	(1.2,2.6)	(-0.7,2.9)	(0.5,3.3)	(0.99,1.02)	(1.99,2.01)	(0.8,1.3)	(1.8,2.2)
200	T.V	1	2	1	2	1	2	1	2	1	2
	Mean	1.26	1.99	1.00	1.99	1.06	1.96	1.01	2.00	1.04	1.91
	S.D.	0.86	0.52	1.29	0.75	0.82	0.42	0.01	0.01	0.07	0.06
	95%CI	(-0.5,2.8)	(0.9,3.0)	(-1.3,3.5)	(0.6,3.3)	(-0.4,2.6)	(1.2,2.7)	(0.99,1.02)	(1.99,2.01)	(0.92,1.19)	(1.78,2.03)

boa (Smith (2007)). The estimates are posterior means using 10,000 iterations of the MCMC sampler (after 10,000 burn-in iterations).

The estimates for the NBMR model were obtained by fitting a truncated Dirichlet Process (DP) mixture model, which leads to a computationally straightforward approximation and can be easily implemented in the freely available WinBUGS software. Two parallel chains of equal length with different initial values were run for the model. The results were based on 10,000 iterations which followed a burn-in period of 40,000 for each chain.

Table 1 compares the posterior means with the true values of  $\beta_0$  and  $\beta_1$  and also gives standard deviations and 95% credible intervals for each of the models considered in this example.

The results of the analysis indicate that the PBRM works well, as all the absolute biases for the estimated parameters turn out to be in the range [0.01, 0.26]. Furthermore, under both ELBMR and NBRM, the true values for both  $\beta_0$  and  $\beta_1$  are recovered successfully indicating that the methods also work well. However, it should be noted that the standard deviations for both parameters are smaller than in the PBMR, giving shorter confidence intervals.

The MCMC sampler for the regression parameters  $\hat{\beta}_0$  and  $\hat{\beta}_1$  can be used to obtain the empirical samples from the joint posterior distributions of the PBMR parameters. These samples can be used to obtain a consistent estimator of the covariance or correlation structure of the parameter estimators, which is difficult to estimate under the classical approach. For example in case (a),

with sample size  $n=100$ , we have

$$\widehat{Cov}\begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{pmatrix} = \begin{pmatrix} 3 & -1 \\ -1 & 6 \end{pmatrix}.$$

### 3.2. Simulation Example 2

In this section we present the results of a second simulation example with the aim of comparing the performance of our approach with the classical mode regression approach. Specifically, we replicate the simulation study in [Kemp and Santos Silva \(2012\)](#), but only for a sample of size 250, and compare their results with the results obtained under our Bayesian mode regression approach.

Simulation data are generated by the simple linear model

$$y_i = \beta_0 + \beta_1 x_i + (1 + vx_i)\epsilon_i, \quad (3.2)$$

where  $x_i$  are generated from a  $\chi_{(3)}^2$  distribution, scaled to have variance 1, and  $\epsilon_i$  are generated as independent draws from a re-scaled log-gamma random variable,

$$\epsilon_i = -\lambda \ln(Z_i), \quad (3.3)$$

where  $Z$  follows a gamma distribution with mean 1 and scale parameter  $\frac{1}{\alpha}$ , to ensure that  $\epsilon_i$  has zero mode. Furthermore, we set  $\lambda = [(1 + 2E(x_i)v + E(x_i^2)v^2)\psi(\alpha)]^{-1}$  to ensure that the unconditional variance of the error  $(1 + vx_i)$  is equal to one.

The study was performed for  $\alpha \in \{0.05, 5\}$  and for  $v \in \{0, 2\}$ . [Table 1](#) compares the 95% Bayesian credible intervals (BCI) for the estimates obtained under PBMR and NBMR with the 95% classical confidence intervals (CI) for the estimates under the two classical mode regression models: Mode 1.6 and Mode 0.8. Mode 1.6 and Mode 0.8 correspond to  $k = 1.6$  and  $k = 0.8$  respectively in the bandwidth selection rule,  $\text{bandwidth} = k \text{mad} n^{-0.143}$ , where  $\text{mad}$  is the median of the absolute deviation from the median of ordinary least squares regression residuals.

The results of the analysis suggest that the Bayesian mode regression estimates are strong competitors of the classical mode regression estimates since in almost all the examples both PBMR

---

<sup>1</sup> $\psi(\cdot)$  is the trigamma function

Table 2: Simulation Example 2: Comparison between Classical and Bayesian approach for mode regression

			PBMR	NBMR	Mode 1.6	Mode 0.8
$\alpha$	n		95% BCI	95% BCI	95% CI	95% CI
5.00	0	$\beta_0$	(-0.37,0.29)	(-0.21,0.36)	(-0.31, 0.41)	(-0.69, 0.75)
		$\beta_1$	(0.82,1.28)	(0.89,1.32)	(0.77, 1.24)	(0.56,1.45)
	2	$\beta_0$	(-0.06,0.07)	(-0.03,0.21)	(-0.15,0.23)	(-0.25,0.29)
		$\beta_1$	(0.99,1.14)	(0.80,1.22)	(0.63,1.37)	(0.48,1.53)
0.05	0	$\beta_0$	(0.00, 0.14)	(-0.03,0.07)	(0.12,0.42)	(-0.09,0.35)
		$\beta_1$	(0.95,1.13)	(0.95,1.06)	(0.90,1.11)	(0.87,1.17)
	2	$\beta_0$	(0.02,0.08)	(0.04,0.09)	(0.09,0.29)	(0.01,0.21)
		$\beta_1$	(0.99,1.08)	(0.97,1.04)	(0.91,1.19)	(0.85,1.19)

and NBMR estimators outperform the two classical estimators.

Finally, as also evident from [Kemp and Santos Silva \(2012\)](#), the selection of the value/prior for  $\sigma$  plays an important role on the precision of the parameters.

### 3.3. Factors Affecting the Body Mass Index (BMI)

Following the introduction of the BMI example in Section 1, the proposed methodology was applied to investigate the research question: “What is the effect of factors such as gender, age, consumption of alcohol, consumption of fruit and vegetables and smoking on the typical body mass index (BMI)?”

A person’s typical BMI was modelled as a function of the person’s age,  $age_i$ , the total units of alcohol consumed per week,  $alcohol_i$ , the portion of fruit and vegetables consumed the previous day,  $fruit\&veg_i$  the person’s cigarette smoking status,  $smoking_i$  (1= Non-smoker, 2= Light smokers, under 10 a day, 3= Moderate smokers, 10 to under 20 a day, 4=Heavy smokers, 20 or more a day), and of a gender indicator,  $male_i$  (1=male, 0=female):

$$bmi_i = \beta_0 + \beta_1 age_i + \beta_2 alcohol_i + \beta_3 fruit\&veg_i + \beta_4 smoking_i + \beta_5 gender_i + \epsilon_i \quad (3.4)$$

The BMI range is from 15.9 to 56.0 (range =40.1) indicating a significant disparity between high and low BMI scores. The average BMI is 27.75 with standard deviation of 5.13. The high



1  
2  
3 levels for range and standard deviation suggest the presence of outliers which cause the mean to be  
4 pulled in the direction of the tail. As a consequence, the mean, median, and mode do not coincide  
5 and it can be easily concluded that the distribution of the data is positively skewed. Figure ??  
6  
7 in section 1 demonstrates the density of BMI for the total, males and females, verifying that all  
8 three distributions are positively skewed. The mode represents the most typical value and is the  
9 value at the peak of the distribution. Even though, mean regression and quantile regression could  
10 have been applied to model BMI these methods cannot reveal any information about the mode, or  
11 about the effect of the covariates on the most typical case.  
12  
13  
14  
15  
16  
17

18 Table 3 presents the estimation results obtained with the traditional mean, quantile regressions  
19 and with the proposed mode regression. The analysis was performed for the total of responders  
20 but also for males and females separately. For the mode regression, an independent improper uni-  
21 form prior was chosen for all the components of  $\beta$  and a gamma prior with mean  $3sd(\mathbf{bmi})$  for  
22  $\sigma$ . Realisations were simulated from the posterior distributions by means of a single-component  
23 Metropolis-Hastings algorithm. Each of the parameters was updated using a random-walk Metropo-  
24 lis algorithm with a Gaussian proposal density centred at the current state of the chain. The  
25 estimates are posterior means using 10,000 iterations of the MCMC sampler (after 10,000 burn-in  
26 iterations).  
27  
28  
29  
30  
31  
32  
33  
34

35 As expected, mean regression indicates that on average the BMI is lower for women than for men  
36 but, as indicated by quantile regression, the effect of gender differs significantly at different quantile  
37 levels. More specifically, at the 25% level, the BMI of women is around 1.36 units lower than the  
38 corresponding BMI for men but this gap is smaller for the median case (0.75) and decreases further  
39 at the 75% quantile level (0.29). Mode regression reveals that the gender differential in the most  
40 typical BMI is lower than both the mean and the median, since as opposed to the other statistics,  
41 mode is not influenced by the extreme observations. According to the results, the typical BMI for  
42 women is 0.27 units lower than the corresponding BMI for men.  
43  
44  
45  
46  
47  
48  
49

50 The effect of fruit and vegetables is not significantly different from zero for mean and quantile  
51 regression at the 25% level and at the median. However, at the 75% level the consumption of  
52 additional fruit and vegetables has a negative effect on the typical BMI (-0.11). Similar results are  
53 obtained for both males and females, however, under quantile regression for males the effect is not  
54 significant at any quantile levels. However, under mode regression it seems that the consumption of  
55  
56  
57  
58  
59  
60

Table 3: BMI dataset: Estimation results for mean, quantile and mode regression

Variable	Mean		Quantile Regression						Parametric Bayesian	
	Regression		0.25		0.50		0.75		Mode Regression	
Total (n=4,138)										
Variable	Estimate	S.E.	Estimate	S.E.	Estimate	S.E.	Estimate	S.E.	Estimate	S.D.
const	25.74	0.33	21.32	0.29	24.04	0.37	28.42	0.53	34.44	10.6
age	0.05	0.004	0.05	0.004	0.06	0.005	0.05	0.01	-0.01	0.16
alcohol	-0.01	0.004	0.004	0.001	0.001	0.004	-0.004	0.001	-0.01	0.05
fruit&veg	-0.07	0.03	-0.002	0.02	-0.06	0.02	-0.11	0.04	-0.24	0.54
smoking	-0.37	0.13	-0.40	0.11	-0.34	0.17	-0.36	0.20	-1.16	3.34
male	0.42	0.16	1.36	0.15	0.75	0.17	0.29	0.24	0.27	6.15
Males (n=1,814)										
const	25.78	0.45	22.81	0.44	24.85	0.47	27.99	0.70	34.7	9.70
age	0.05	0.006	0.05	0.006	0.06	0.01	0.05	0.01	-0.003	0.15
alcohol	0.01	0.004	-0.01	0.005	0.01	0.01	0.01	0.01	-0.01	0.06
fruit&veg	-0.03	0.04	-0.02	0.04	-0.03	0.04	-0.02	0.06	-0.31	0.68
smoking	-0.43	0.19	-0.57	0.20	-0.48	0.24	-0.51	0.33	-1.47	3.53
Females (n=2,324)										
const	26.05	0.47	21.36	0.40	23.76	0.47	28.51	0.83	33.8	10.25
age	0.04	0.006	0.05	0.006	0.06	0.01	0.05	0.01	0.01	0.16
alcohol	-0.03	0.01	-0.01	0.01	-0.01	0.01	-0.02	0.01	-0.04	0.01
fruit&veg	-0.1	0.04	0.01	0.03	-0.07	0.04	-0.15	0.06	-0.17	0.53
smoking	-0.27	0.20	-0.29	0.18	-0.19	0.27	-0.29	0.37	-1.32	3.73

an additional unit of fruit and vegetables is negatively correlated with the typical BMI level. The negative effect for males is almost twice as high as for females. This results imply that, typically, eating more fruit and vegetables can contribute to lowering BMI levels; thus losing weight.

Furthermore, the results of the analysis suggest that heavier smoking is also negatively correlated with the BMI under all three methods. However, under mode regression the negative effect of heavier smoking on the typical BMI is 3 times higher as compared to the effect on the mean BMI and at different quantile levels.

Finally, under all three methods, the effect of age and alcohol cannot be considered as significantly different from zero for the total, but also for males and females separately.

In conclusion, the results indicate that mode regression is a useful statistical technique, especially when analysing data with outliers. In this example, even though the overall effect of covariates on the response variable was similar under the three regression methods, the marginal effects of the covariates were often different, justifying the usefulness of mode regression as an alternative analysis tool.

#### 4. CONCLUSIONS

Identifying the typical value or pattern could be one of the most efficient statistical methods of data analysis, in particular, for big data analysis. In this paper a novel Bayesian mode regression framework has been presented which includes three approaches: a parametric method, a nonparametric method and an empirical likelihood-based method. It should be noted that, in the area of mode regression, there is no literature from a Bayesian perspective. The paper demonstrates that the estimates are consistent and asymptotically Normal under fairly standard conditions and even under misspecification of the likelihood function. The numerical studies suggest that the proposed Bayesian mode regression estimates are strong competitors to the classical mode regression estimates.

#### APPENDIX: PROOFS OF THEOREMS

##### Proof of theorem 2.1

The  $\gamma$ th moments of marginal posterior distribution of  $\beta$  is given by

$$E[|\beta|^\gamma | \sigma, \mathbf{y}] = \int \frac{1}{(2\sigma)^n} \prod_{i=1}^n I[|y_i - x'_i \beta| < \sigma] \pi(\beta, \sigma) d\beta d\sigma.$$

Noting that  $\prod_{i=1}^n I[|y_i - x'_i \beta| < \sigma]$  provides joint bands for all components  $\beta_j$  ( $j = 0, 1, \dots, p$ ) of  $\beta$ . Let us say  $0 < |\beta_j| < B_j < \infty$  ( $j = 0, 1, \dots, p$ ), even if some of  $|y_i - x'_i \beta| < \sigma$  are true and some are not. Therefore,

$$E[|\beta|^\gamma | \sigma, \mathbf{y}] = \int \frac{1}{(2\sigma)^n} d\sigma \int_{-B_0}^{B_0} \int_{-B_1}^{B_1} \dots \int_{-B_p}^{B_p} \prod_{j=0}^p |\beta_j|^\gamma \pi(\beta, \sigma) d\beta,$$

which is clearly finite. Similarly, for the  $\gamma$ th moment of marginal posterior of  $\sigma$  with  $\gamma < n$  is defined as  $E[|\sigma|^\gamma | \beta, \mathbf{y}]$ , and can be provided finite in the same way.

##### Proof of theorem 2.2

We will show Theorem 2.2 by applying a generic consistency lemma, Lemma 4.1, of Lu et al. (2007). For convenience of statement, we define  $R_n(\lambda, \beta) \equiv n^{-1} \sum_{i=1}^n \log(1 + \lambda' g(X_i, Y_i, \beta))$  and  $R(\lambda, \beta) \equiv E\{\log(1 + \lambda' g(X_i, Y_i, \beta))\}$ . Then note that  $R_n(\hat{\lambda}(\beta), \beta) = -\Gamma_n(\beta)$  and  $\hat{\beta} = \arg \min_{\beta \in \mathbb{B}} R_n(\hat{\lambda}(\beta), \beta)$ , where  $\hat{\lambda}(\beta)$  and  $\Gamma_n(\beta)$  are defined in (2.13) and (2.14), respectively, and  $\mathbb{B}$  is a compact subset of  $\mathbb{R}^p$  containing the true parameter vector  $\beta_0$  as an interior point. Further, we denote by  $H_n(\lambda, \beta)$  for the left-hand side of (2.13) divided by  $n$ , that is  $H_n(\lambda, \beta) \equiv n^{-1} \sum_{i=1}^n \{g(X_i, Y_i, \beta) / [1 + \lambda' g(X_i, Y_i, \beta)]\}$ , and hence for any  $\beta \in \mathbb{B}$ ,  $\hat{\lambda}(\beta)$

is the solution of  $\lambda$  to the equation  $H_n(\lambda, \beta) = 0$ .

We will need the following lemma on the continuity for the quantities related.

**Lemma .1.** *Under Assumptions 2 and 3, we have the following results:*

(L1)  $E\{g(X, Y, \beta)\}$  and  $E\{g(X, Y, \beta)g(X, Y, \beta)'\}$  are twice continuously differentiable with respect to  $\beta$ .

(L2) There exist  $p$  dimensional compact neighborhoods  $C_\lambda$  and  $C_\beta$  around 0, in which  $H_0(\lambda, \beta) = E[g(X, Y, \beta)/\{1+\lambda'g(X, Y, \beta)\}]$  is twice continuously differentiable in  $\beta \in C_\beta$  and  $\lambda \in C_\lambda$ , and  $E[g(X, Y, \beta)g(X, Y, \beta)'/\{1+\lambda'g(X, Y, \beta)\}]$  is uniformly continuous with respect to  $\beta \in C_\beta$  and  $\lambda \in C_\lambda$ .

The proof of this lemma is similar to that of Lemma A.1 of Yang and He (2012, pp. 1121). We only need to notice  $g(X_i, Y_i, \beta) = (Y_i - \beta'X_i)I_{\{|Y_i - \beta'X_i| < \sigma\}}$  and apply Assumptions 2 and 3. As an illustration, we provide the proof for  $Eg(X_i, Y_i, \beta)$  here. Note that

$$\begin{aligned} Eg(X_i, Y_i, \beta) &= E_X \int (y - \beta'X) I_{\{|y - \beta'X| < \sigma\}} X f_X(y) dy \\ &= E_X \int_{\beta'X - \sigma}^{\beta'X + \sigma} (y - \beta'X) X f_X(y) dy, \end{aligned}$$

where  $E_X$  stands for the expectation with respect to the distribution  $G_X$  of the random variable  $X$ . Then the first order derivative of  $Eg(X_i, Y_i, \beta)$  with respect to  $\beta$ , through simple algebraic calculations, is

$$\frac{\partial Eg(X_i, Y_i, \beta)}{\partial \beta} = E_X \{ \sigma X (f_X(\beta'X + \sigma) - f_X(\beta'X - \sigma)) - XX'(F_X(\beta'X + \sigma) - F_X(\beta'X - \sigma)) \}.$$

Now by Assumptions 2 and 3, clearly  $\frac{\partial Eg(X_i, Y_i, \beta)}{\partial \beta}$  is further differentiable with respect to  $\beta$ . The remaining parts of this lemma can be proved similarly with details omitted. ‡

We further define  $\lambda_0(\beta)$  to be the solution of  $\lambda$  to the equation  $H(\lambda, \beta) \equiv E\{g(X_i, Y_i, \beta)/[1+\lambda'g(X_i, Y_i, \beta)]\} = 0$ . By Lemma .1, Assumption 5 and the implicit function theorem,  $\lambda_0(\beta)$  uniquely exists in the neighbourhood  $C_\lambda$  of  $\mathbf{0} \in \mathbb{R}^p$ . By this uniqueness, as  $Eg(X, Y, \beta_0) = 0$ , we have  $\lambda_0(\beta_0) = 0$ . Therefore it follows that  $R(\lambda_0(\beta_0), \beta_0) = E\{\log(1 + (\lambda_0(\beta_0))'g(X_i, Y_i, \beta_0))\} = 0$ . Note that under Assumptions 1–5,  $\beta_0 = \arg \min_{\beta \in \mathbb{B}} R(\lambda_0(\beta), \beta)$ .

To show the consistency of  $\hat{\beta}$  to  $\beta_0$ , we will apply a lemma below that is a special case of Lemma 4.1 of Lu et al. (2007). Here we need to define a uniform metric  $\|\cdot\|_{\mathbb{B}}$  for the distance of any continuous function  $\lambda: \mathbb{B} \mapsto \mathbb{R}^p$  from  $\lambda_0(\cdot)$ , that is  $\|\lambda(\cdot) - \lambda_0(\cdot)\|_{\mathbb{B}} = \sup_{\beta \in \mathbb{B}} \|\lambda(\beta) - \lambda_0(\beta)\|$  with  $\|\cdot\|$  standing for the Euclidean norm of  $\mathbb{R}^p$ .

**Lemma .2.** *Suppose  $\beta_0 \in \mathbb{B}$  (a compact subset of  $\mathbb{R}^p$ ) satisfies  $R(\lambda_0(\beta_0), \beta_0) = \inf_{\beta \in \mathbb{B}} R(\lambda_0(\beta), \beta)$ , and that the following hold.*

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

$$(i) R_n(\hat{\lambda}(\hat{\beta}), \hat{\beta}) \leq \inf_{\beta \in \mathbb{B}} R_n(\hat{\lambda}(\beta), \beta) + o_P(1).$$

(ii) For all  $\delta > 0$ , there exists  $\epsilon(\delta) > 0$  such that

$$\inf_{\|\beta - \beta_0\| > \delta} R(\lambda_0(\beta), \beta) \geq R(\lambda_0(\beta_0), \beta_0) + \epsilon(\delta).$$

(iii) Uniformly for all  $\beta \in \mathbb{B}$ ,  $R(\lambda(\beta), \beta)$  is continuous [with respect to the uniform metric  $\|\cdot\|_{\mathbb{B}}$ ] in  $\lambda(\beta)$  at  $\lambda_0(\beta)$ .

(iv)  $\|\hat{\lambda}(\cdot) - \lambda_0(\cdot)\|_{\mathbb{B}} = o_P(1)$ .

(v) For all  $\{\delta_n\}$  with  $\delta_n = o(1)$ ,

$$\sup_{\beta \in \mathbb{B}} \sup_{\|\lambda(\beta) - \lambda_0(\beta)\|_{\mathbb{B}} \leq \delta_n} |R_n(\lambda(\beta), \beta) - R(\lambda(\beta), \beta)| = o_P(1).$$

Then  $\hat{\beta} - \beta_0 = o_P(1)$ .

The proof of this lemma is omitted; see that of Lemma 4.1 of [Lu et al. \(2007, pp. 186\)](#).

The consistency of  $\hat{\beta}$  can be proved by checking the conditions in Lemma .2 step by step: As  $\hat{\beta}$  and  $\beta_0$  are the minimizers of  $R_n(\hat{\lambda}(\beta), \beta)$  and  $R(\lambda_0(\beta), \beta)$ , respectively, (i) and (ii) hold obviously. By noting Lemma .1, simple algebraic calculations lead to

$$R(\lambda, \beta) = E_X \int_{\beta'X - \sigma}^{\beta'X + \sigma} \log\{1 + \lambda'X(y - \beta'X)\} f_X(y) dy, \tag{.1}$$

$$H(\lambda_0(\beta), \beta) = E_X \int_{\beta'X - \sigma}^{\beta'X + \sigma} \frac{X(y - \beta'X)}{1 + (\lambda_0(\beta))'X(y - \beta'X)} f_X(y) dy = 0, \tag{.2}$$

and therefore (iii) also holds clearly by the following fact: as  $\|\lambda(\cdot) - \lambda_0(\cdot)\|_{\mathbb{B}} \rightarrow 0$ ,

$$\begin{aligned}
& \sup_{\boldsymbol{\beta} \in \mathbb{B}} |R(\lambda(\boldsymbol{\beta}), \boldsymbol{\beta}) - R(\lambda_0(\boldsymbol{\beta}), \boldsymbol{\beta})| \\
& \leq \sup_{\boldsymbol{\beta} \in \mathbb{B}} \left| E_X \int_{\boldsymbol{\beta}'X-\sigma}^{\boldsymbol{\beta}'X+\sigma} [\log\{1 + (\lambda(\boldsymbol{\beta}))'X(y - \boldsymbol{\beta}'X)\} - \log\{1 + (\lambda_0(\boldsymbol{\beta}))'X(y - \boldsymbol{\beta}'X)\}] f_X(y) dy \right| \\
& \leq \sup_{\boldsymbol{\beta} \in \mathbb{B}} \left| E_X \int_{\boldsymbol{\beta}'X-\sigma}^{\boldsymbol{\beta}'X+\sigma} \left[ \frac{(\lambda(\boldsymbol{\beta}) - \lambda_0(\boldsymbol{\beta}))'X(y - \boldsymbol{\beta}'X)}{1 + (\lambda_0(\boldsymbol{\beta}))'X(y - \boldsymbol{\beta}'X)} \right. \right. \\
& \quad \left. \left. - \frac{(\lambda(\boldsymbol{\beta}) - \lambda_0(\boldsymbol{\beta}))'XX'(y - \boldsymbol{\beta}'X)^2(\lambda(\boldsymbol{\beta}) - \lambda_0(\boldsymbol{\beta}))}{[1 + (\lambda_0(\boldsymbol{\beta}) + \xi(\lambda(\boldsymbol{\beta}) - \lambda_0(\boldsymbol{\beta}))]'X(y - \boldsymbol{\beta}'X)]^2} \right] f_X(y) dy \right| \\
& \leq \|\lambda(\cdot) - \lambda_0(\cdot)\|_{\mathbb{B}}^2 \sup_{\boldsymbol{\beta} \in \mathbb{B}} \left| E_X \int_{\boldsymbol{\beta}'X-\sigma}^{\boldsymbol{\beta}'X+\sigma} \left[ \frac{\|XX'\|(y - \boldsymbol{\beta}'X)^2}{[1 + (\lambda_0(\boldsymbol{\beta}))'X(y - \boldsymbol{\beta}'X)]^2} \right] f_X(y) dy \right| \rightarrow 0, \tag{.3}
\end{aligned}$$

where  $|\xi| < 1$ , the last inequality follows from equality of (.2), and the last limit is owing to the compactness of  $\mathbb{B}$  together with the continuity of the integration part as a function of  $\boldsymbol{\beta}$  on the RHS of the last inequality in (.3). (iv) follows from a standard argument of the Z-estimator  $\hat{\lambda}(\boldsymbol{\beta})$ , which is the solution to  $H_n(\lambda, \boldsymbol{\beta}) = 0$ , uniformly converging to  $\lambda_0(\boldsymbol{\beta})$ , which is the solution to  $H(\lambda, \boldsymbol{\beta}) = 0$ , in Chapter 5.1 of [Van der Vaart \(1998\)](#); see also the argument on uniform convergence in the second paragraph on [Yang and He \(2012, pp. 1124\)](#). For (v), letting  $\delta_n = o(1)$  and  $\|\lambda - \lambda_0\|_{\mathbb{B}} \leq \delta_n$ , we notice that

$$\begin{aligned}
& R_n(\lambda(\boldsymbol{\beta}), \boldsymbol{\beta}) - R(\lambda(\boldsymbol{\beta}), \boldsymbol{\beta}) \\
& = \{R_n(\lambda(\boldsymbol{\beta}), \boldsymbol{\beta}) - R_n(\lambda_0(\boldsymbol{\beta}), \boldsymbol{\beta})\} + \{R_n(\lambda_0(\boldsymbol{\beta}), \boldsymbol{\beta}) - R(\lambda_0(\boldsymbol{\beta}), \boldsymbol{\beta})\} \\
& \quad + \{R(\lambda_0(\boldsymbol{\beta}), \boldsymbol{\beta}) - R(\lambda(\boldsymbol{\beta}), \boldsymbol{\beta})\} \\
& = I + II + III,
\end{aligned}$$

where by (.3)  $III$  tends to 0, uniformly for  $\boldsymbol{\beta} \in \mathbb{B}$  and with  $\lambda$  satisfying  $\|\lambda - \lambda_0\|_{\mathbb{B}} \leq \delta_n$ . That  $I$  tends to 0, uniformly for  $\boldsymbol{\beta} \in \mathbb{B}$  and  $\lambda$  with  $\|\lambda - \lambda_0\|_{\mathbb{B}} \leq \delta_n$ , can be proved in the same way as for  $III$ , because in fact  $E[I] = III$ ;  $II$  can also be proved easily to tend to zero.

### Proof of theorem 2.3

Based on the consistency in Theorem 2.2, Theorem 2.3 can be proved similarly to Theorem 3.2 of [Yang and He \(2012\)](#) by noticing the difference of mode regression in this paper from quantile regression in [Yang and He \(2012\)](#). First, under Assumptions 2–4, it is easy to show as done in Lemma A.5 of [Yang and He \(2012\)](#) that

$$(C1) \left\| \sum_{i=1}^n [g(X_i, Y_i, \boldsymbol{\beta}) - Eg(X_i, Y_i, \boldsymbol{\beta})] \right\| = O_p(n^{1/2}), \text{ uniformly in } \boldsymbol{\beta} \text{ in a } o(1)\text{-neighborhood of } \boldsymbol{\beta}_0.$$

$$(C2) \left\| \sum_{i=1}^n [g(X_i, Y_i, \boldsymbol{\beta})g(X_i, Y_i, \boldsymbol{\beta})' - Eg(X_i, Y_i, \boldsymbol{\beta})g(X_i, Y_i, \boldsymbol{\beta})'] \right\| = o_p(n), \text{ uniformly in } \boldsymbol{\beta} \text{ in a } o(1)\text{-}$$

neighborhood of  $\beta_0$ .

(C3)  $\|\sum_{i=1}^n [g(X_i, Y_i, \beta) - Eg(X_i, Y_i, \beta) - g(X_i, Y_i, \beta_0) + Eg(X_i, Y_i, \beta_0)]\| = o_p(n^{-1/2})$ , uniformly in  $\beta$  for  $\beta - \beta_0 = O_p(n^{-1/2})$ .

These (C1)-(C3) together with Assumptions 1-5 ensure (2.15) holds true (c.f., Lemma 6 of Molanes Lopez et al. (2009)).

Further, maximizing the main terms on the RHS of (2.15) with respect to  $\beta$ , we have

$$\hat{\beta} - \beta_0 = n^{-1/2}(V'_{12}V^{-1}_{11}V_{12})^{-1}V'_{12}V^{-1}_{11}W_n + o_P(n^{-1/2}), \quad (.4)$$

where  $\hat{\beta}$  is the maximum empirical likelihood estimator of  $\beta_0$ .

Then it follows from (2.15) and (.4) that

$$\begin{aligned} \pi(\beta|data) &= \pi(\beta) \mathfrak{R}(\beta) \\ &= \pi(\beta) \times \exp \left\{ -\frac{n}{2}(\beta - \beta_0)'V'_{12}V^{-1}_{11}V_{12}(\beta - \beta_0) + n^{1/2}(\beta - \beta_0)'V'_{12}V^{-1}_{11}W_n \right. \\ &\quad \left. - \frac{1}{2}W'_nV^{-1}_{11}W_n + o_P(1) \right\} \\ &= \pi(\beta) \times \exp \left\{ -\frac{n}{2}(\beta - \beta_0)'V'_{12}V^{-1}_{11}V_{12}(\beta - \beta_0) + n(\beta - \beta_0)'V'_{12}V^{-1}_{11}V_{12}(\hat{\beta} - \beta_0) \right. \\ &\quad \left. - \frac{1}{2}W'_nV^{-1}_{11}W_n + o_P(1) \right\} \\ &= \pi(\beta) \times \exp \left\{ -\frac{n}{2}(\beta - \beta_0)'V'_{12}V^{-1}_{11}V_{12}(\beta - 2\hat{\beta} + \beta_0) - \frac{1}{2}W'_nV^{-1}_{11}W_n + o_P(1) \right\} \\ &= \pi(\beta) \exp \left\{ -\frac{n}{2}(\beta - \hat{\beta})'I_n(\beta - \hat{\beta}) + Q_n \right\}, \end{aligned} \quad (.5)$$

where, by (.4),

$$\begin{aligned} Q_n &= -\frac{n}{2}(\hat{\beta} - \beta_0)'V'_{12}V^{-1}_{11}V_{12}(\beta - 2\hat{\beta} + \beta_0) + \frac{n}{2}(\beta - \hat{\beta})'V'_{12}V^{-1}_{11}V_{12}(\hat{\beta} - \beta_0) \\ &\quad - \frac{1}{2}W'_nV^{-1}_{11}W_n + o_P(1) \\ &= \frac{n}{2}(\hat{\beta} - \beta_0)'V'_{12}V^{-1}_{11}V_{12}(\hat{\beta} - \beta_0) - \frac{1}{2}W'_nV^{-1}_{11}W_n + o_P(1) \\ &= \frac{n}{2}(n^{-1/2}(V'_{12}V^{-1}_{11}V_{12})^{-1}V'_{12}V^{-1}_{11}W_n + o_P(n^{-1/2}))'V'_{12}V^{-1}_{11}V_{12} \\ &\quad \times (n^{-1/2}(V'_{12}V^{-1}_{11}V_{12})^{-1}V'_{12}V^{-1}_{11}W_n + o_P(n^{-1/2})) - \frac{1}{2}W'_nV^{-1}_{11}W_n + o_P(1) \\ &= \frac{1}{2}W'_nV^{-1}_{11}W_n + o_P(1) - \frac{1}{2}W'_nV^{-1}_{11}W_n + o_P(1) = o_P(1). \end{aligned} \quad (.6)$$

Therefore (2.16) follows from (.5) and (.6) together with  $\log(\pi(\beta)) = \log(\pi(\beta_0)) + O(n^{-1/2})$  for  $\beta - \beta_0 = O(n^{-1/2})$  owing to Assumption 6.

1  
2  
3 The remaining part of Theorem 2.3 can be proved, by using Assumption 6, as done in the corresponding  
4 proof of Theorem 3.2 of Lu et al. (2007, pp. 186). The details are therefore omitted.  
5  
6  
7

## 8 REFERENCES

- 9  
10  
11 Berlinet, A., Vajda, I., and Van der Meulen, E. (1998). About the asymptotic accuracy of barron  
12 density estimates. *Information Theory, IEEE Transactions*, 44(3):999–1009.  
13  
14  
15  
16 Bickel, P. and Fan, J. (1996). Some problems on the estimation of unimodal densities. *Statistica*  
17 *Sinica*, 6:23–46.  
18  
19  
20 Birgé, L. (1997). Estimation of unimodal densities without smoothness assumptions. *The Annals*  
21 *of Statistics*, 25(3):970–981.  
22  
23  
24  
25 Brunner, L. (1992). Bayesian nonparametric methods for data from a unimodal density. *Statistics*  
26 *& Probability letters*, 14(3):195–199.  
27  
28  
29 Chernozhukov, V. and Hong, H. (2003). An mcmc approach to classical estimation. *Journal of*  
30 *Econometrics*, 115(2):293–346.  
31  
32  
33  
34 Dunson, D., Pillai, N., and Park, J. (2007). Bayesian density regression. *Journal of the Royal*  
35 *Statistical Society: Series B (Statistical Methodology)*, 69(2):163–183.  
36  
37  
38 Eddy, W. (1980). Optimum kernel estimators of the mode. *The Annals of Statistics*, 8(4):870–882.  
39  
40  
41 Feller, W. (1971). *An introduction to probability theory and its applications*, volume 2. Wiley-New  
42 york.  
43  
44  
45 Gasser, T., Hall, P., and Presnell, B. (1998). Nonparametric estimation of the mode of a distribution  
46 of random curves. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*,  
47 60(4):681–691.  
48  
49  
50  
51 Grenander, U. (1965). Some direct estimates of the mode. *The Annals of Mathematical Statistics*,  
52 36(1):131–138.  
53  
54  
55  
56 Hall, P. and Huang, L. (2001). Nonparametric kernel regression subject to monotonicity constraints.  
57 *The Annals of Statistics*, 29(3):624–647.  
58  
59  
60

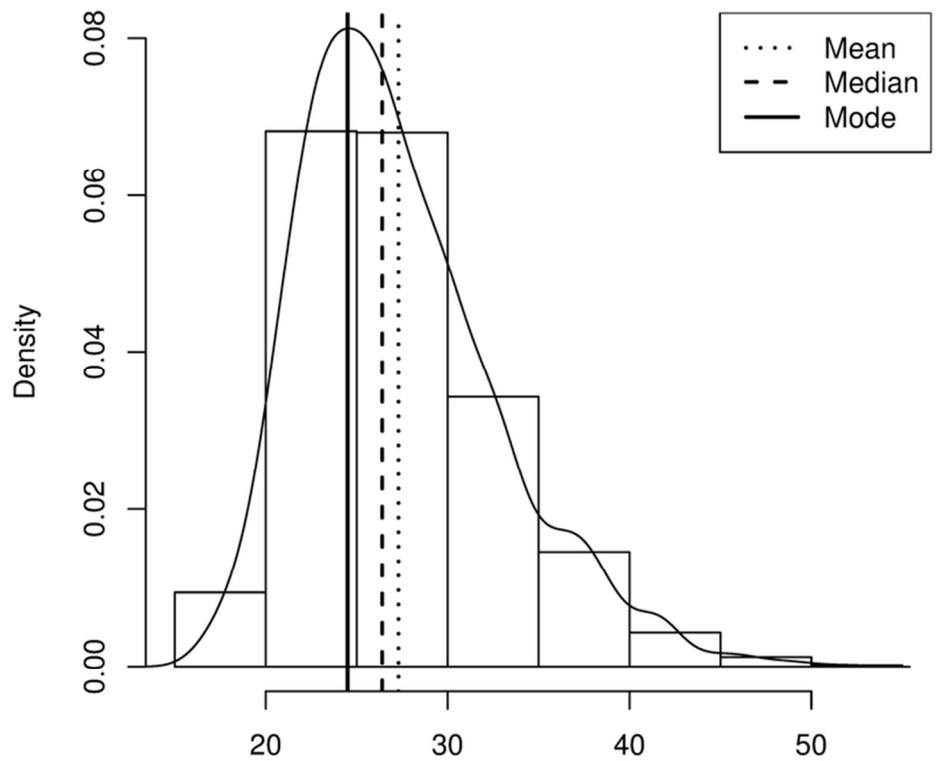


- 1  
2  
3 Hall, P., Peng, L., and Rau, C. (2001). Local likelihood tracking of fault lines and boundaries.  
4  
5 *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63(3):569–582.  
6  
7  
8 Heckman, D., Geiser, D., Eidell, B., Stauffer, R., Kardos, N., and Hedges, S. (2001). Molecular  
9  
10 evidence for the early colonization of land by fungi and plants. *Science*, 293(5532):1129.  
11  
12  
13 Hedges, S. and Shah, P. (2003). Comparison of mode estimation methods and application in  
14  
15 molecular clock analysis. *BMC Bioinformatics*, 4(1):31.  
16  
17  
18 Kemp, G. C. and Santos Silva, J. (2012). Regression towards the mode. *Journal of Econometrics*,  
19  
20 170(1):92–101.  
21  
22 Kumar, S. and Hedges, S. (1998). A molecular timescale for vertebrate evolution. *Nature*,  
23  
24 392(6679):917–920.  
25  
26  
27 Lee, M. (1989). Mode regression. *Journal of Econometrics*, 42(3):337–349.  
28  
29  
30 Lee, M. (1993). Quadratic mode regression. *Journal of Econometrics*, 57(1-3):1–19.  
31  
32  
33 Lu, Z., Tjstheim, D., and Yao, Q. (2007). Adaptive varying-coefficient linear models for stochastic  
34  
35 processes: asymptotic theory. *Statistica Sinica*, 17:177–197.  
36  
37  
38 Manski, C. (1991). Regression. *Journal of Economic Literature*, 29(1):34–50.  
39  
40  
41 Markov, H., Valtchev, T., Borissova, J., and Golev, V. (1997). An algorithm to "clean" close stellar  
42  
43 companions. *Astronomy and Astrophysics Supplement Series*, 122(1):193–199.  
44  
45  
46 Meyer, M. (2001). An alternative unimodal density estimator with a consistent estimate of the  
47  
48 mode. *Statistica Sinica*, 11(4):1159–1174.  
49  
50  
51 Molanes Lopez, E., Keilegom, I., and Veraverbeke, N. (2009). Empirical likelihood for non-smooth  
52  
53 criterion functions. *Scandinavian Journal of Statistics*, 36(3):413–432.  
54  
55  
56 Parzen, E. (1962). On estimation of a probability density function and mode. *The Annals of*  
57  
58 *Mathematical Statistics*, 33(3):1065–1076.  
59  
60  
61  
62  
63  
64  
65  
66  
67  
68  
69  
70  
71  
72  
73  
74  
75  
76  
77  
78  
79  
80  
81  
82  
83  
84  
85  
86  
87  
88  
89  
90  
91  
92  
93  
94  
95  
96  
97  
98  
99  
100  
101  
102  
103  
104  
105  
106  
107  
108  
109  
110  
111  
112  
113  
114  
115  
116  
117  
118  
119  
120  
121  
122  
123  
124  
125  
126  
127  
128  
129  
130  
131  
132  
133  
134  
135  
136  
137  
138  
139  
140  
141  
142  
143  
144  
145  
146  
147  
148  
149  
150  
151  
152  
153  
154  
155  
156  
157  
158  
159  
160  
161  
162  
163  
164  
165  
166  
167  
168  
169  
170  
171  
172  
173  
174  
175  
176  
177  
178  
179  
180  
181  
182  
183  
184  
185  
186  
187  
188  
189  
190  
191  
192  
193  
194  
195  
196  
197  
198  
199  
200  
201  
202  
203  
204  
205  
206  
207  
208  
209  
210  
211  
212  
213  
214  
215  
216  
217  
218  
219  
220  
221  
222  
223  
224  
225  
226  
227  
228  
229  
230  
231  
232  
233  
234  
235  
236  
237  
238  
239  
240  
241  
242  
243  
244  
245  
246  
247  
248  
249  
250  
251  
252  
253  
254  
255  
256  
257  
258  
259  
260  
261  
262  
263  
264  
265  
266  
267  
268  
269  
270  
271  
272  
273  
274  
275  
276  
277  
278  
279  
280  
281  
282  
283  
284  
285  
286  
287  
288  
289  
290  
291  
292  
293  
294  
295  
296  
297  
298  
299  
300  
301  
302  
303  
304  
305  
306  
307  
308  
309  
310  
311  
312  
313  
314  
315  
316  
317  
318  
319  
320  
321  
322  
323  
324  
325  
326  
327  
328  
329  
330  
331  
332  
333  
334  
335  
336  
337  
338  
339  
340  
341  
342  
343  
344  
345  
346  
347  
348  
349  
350  
351  
352  
353  
354  
355  
356  
357  
358  
359  
360  
361  
362  
363  
364  
365  
366  
367  
368  
369  
370  
371  
372  
373  
374  
375  
376  
377  
378  
379  
380  
381  
382  
383  
384  
385  
386  
387  
388  
389  
390  
391  
392  
393  
394  
395  
396  
397  
398  
399  
400  
401  
402  
403  
404  
405  
406  
407  
408  
409  
410  
411  
412  
413  
414  
415  
416  
417  
418  
419  
420  
421  
422  
423  
424  
425  
426  
427  
428  
429  
430  
431  
432  
433  
434  
435  
436  
437  
438  
439  
440  
441  
442  
443  
444  
445  
446  
447  
448  
449  
450  
451  
452  
453  
454  
455  
456  
457  
458  
459  
460  
461  
462  
463  
464  
465  
466  
467  
468  
469  
470  
471  
472  
473  
474  
475  
476  
477  
478  
479  
480  
481  
482  
483  
484  
485  
486  
487  
488  
489  
490  
491  
492  
493  
494  
495  
496  
497  
498  
499  
500  
501  
502  
503  
504  
505  
506  
507  
508  
509  
510  
511  
512  
513  
514  
515  
516  
517  
518  
519  
520  
521  
522  
523  
524  
525  
526  
527  
528  
529  
530  
531  
532  
533  
534  
535  
536  
537  
538  
539  
540  
541  
542  
543  
544  
545  
546  
547  
548  
549  
550  
551  
552  
553  
554  
555  
556  
557  
558  
559  
560  
561  
562  
563  
564  
565  
566  
567  
568  
569  
570  
571  
572  
573  
574  
575  
576  
577  
578  
579  
580  
581  
582  
583  
584  
585  
586  
587  
588  
589  
590  
591  
592  
593  
594  
595  
596  
597  
598  
599  
600  
601  
602  
603  
604  
605  
606  
607  
608  
609  
610  
611  
612  
613  
614  
615  
616  
617  
618  
619  
620  
621  
622  
623  
624  
625  
626  
627  
628  
629  
630  
631  
632  
633  
634  
635  
636  
637  
638  
639  
640  
641  
642  
643  
644  
645  
646  
647  
648  
649  
650  
651  
652  
653  
654  
655  
656  
657  
658  
659  
660  
661  
662  
663  
664  
665  
666  
667  
668  
669  
670  
671  
672  
673  
674  
675  
676  
677  
678  
679  
680  
681  
682  
683  
684  
685  
686  
687  
688  
689  
690  
691  
692  
693  
694  
695  
696  
697  
698  
699  
700  
701  
702  
703  
704  
705  
706  
707  
708  
709  
710  
711  
712  
713  
714  
715  
716  
717  
718  
719  
720  
721  
722  
723  
724  
725  
726  
727  
728  
729  
730  
731  
732  
733  
734  
735  
736  
737  
738  
739  
740  
741  
742  
743  
744  
745  
746  
747  
748  
749  
750  
751  
752  
753  
754  
755  
756  
757  
758  
759  
760  
761  
762  
763  
764  
765  
766  
767  
768  
769  
770  
771  
772  
773  
774  
775  
776  
777  
778  
779  
780  
781  
782  
783  
784  
785  
786  
787  
788  
789  
790  
791  
792  
793  
794  
795  
796  
797  
798  
799  
800  
801  
802  
803  
804  
805  
806  
807  
808  
809  
810  
811  
812  
813  
814  
815  
816  
817  
818  
819  
820  
821  
822  
823  
824  
825  
826  
827  
828  
829  
830  
831  
832  
833  
834  
835  
836  
837  
838  
839  
840  
841  
842  
843  
844  
845  
846  
847  
848  
849  
850  
851  
852  
853  
854  
855  
856  
857  
858  
859  
860  
861  
862  
863  
864  
865  
866  
867  
868  
869  
870  
871  
872  
873  
874  
875  
876  
877  
878  
879  
880  
881  
882  
883  
884  
885  
886  
887  
888  
889  
890  
891  
892  
893  
894  
895  
896  
897  
898  
899  
900  
901  
902  
903  
904  
905  
906  
907  
908  
909  
910  
911  
912  
913  
914  
915  
916  
917  
918  
919  
920  
921  
922  
923  
924  
925  
926  
927  
928  
929  
930  
931  
932  
933  
934  
935  
936  
937  
938  
939  
940  
941  
942  
943  
944  
945  
946  
947  
948  
949  
950  
951  
952  
953  
954  
955  
956  
957  
958  
959  
960  
961  
962  
963  
964  
965  
966  
967  
968  
969  
970  
971  
972  
973  
974  
975  
976  
977  
978  
979  
980  
981  
982  
983  
984  
985  
986  
987  
988  
989  
990  
991  
992  
993  
994  
995  
996  
997  
998  
999  
1000

- 1  
2  
3 Roberts, G. and Rosenthal, J. (2001). Optimal scaling for various metropolis-hastings algorithms.  
4  
5 *Statistical Science*, 16(4):351–367.  
6  
7  
8 Silverman, B. (1986). *Density estimation for statistics and data analysis*, volume 26. Chapman &  
9  
10 Hall/CRC.  
11  
12 Smith, B. (2007). boa: an r package for mcmc output convergence assessment and posterior  
13  
14 inference. *Journal of Statistical Software*, 21(11):1–37.  
15  
16  
17 Van der Vaart, A. (1998). *Asymptotic statistics*. Cambridge Univ Pr.  
18  
19  
20 Yang, Y. and He, X. (2012). Bayesian empirical likelihood for quantile regression. *Annals of*  
21  
22 *Statistics*, 40(12):1102–1131.  
23  
24 Yao, W. and Li, L. (2013). A new regression model: modal linear regression. *Scandinavian Journal*  
25  
26 *of Statistics*.  
27  
28  
29 Yasukawa, K. (1926). On the probable error of the mode of skew frequency distributions.  
30  
31 *Biometrika*, 18(3/4):263–292.  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

### BMI – female

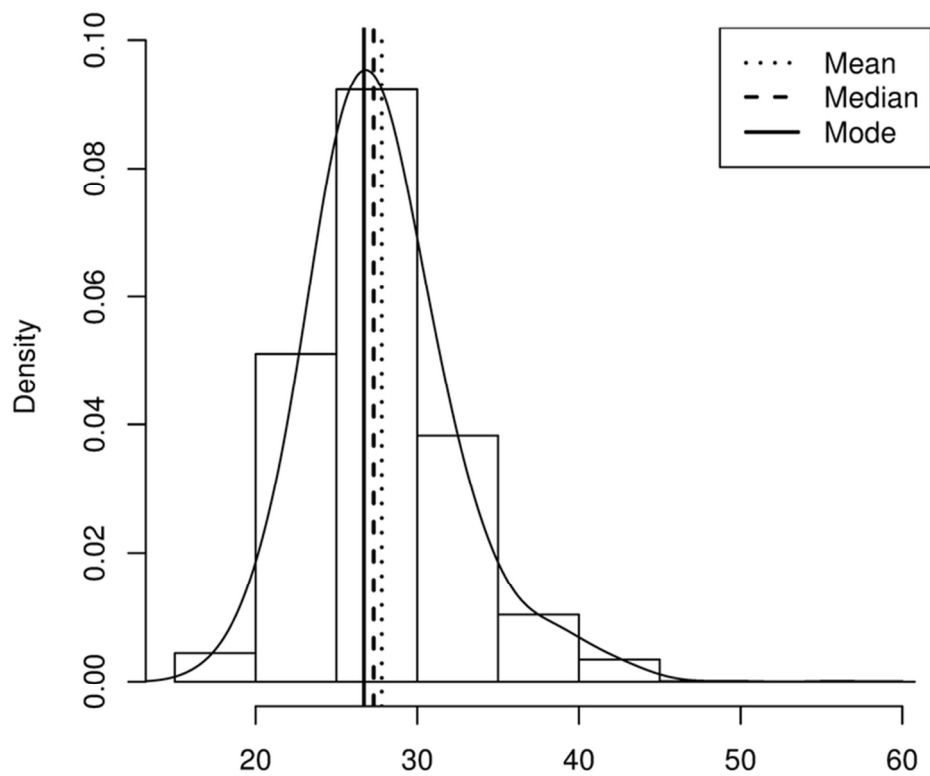


145x145mm (150 x 150 DPI)



1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

### BMI – male

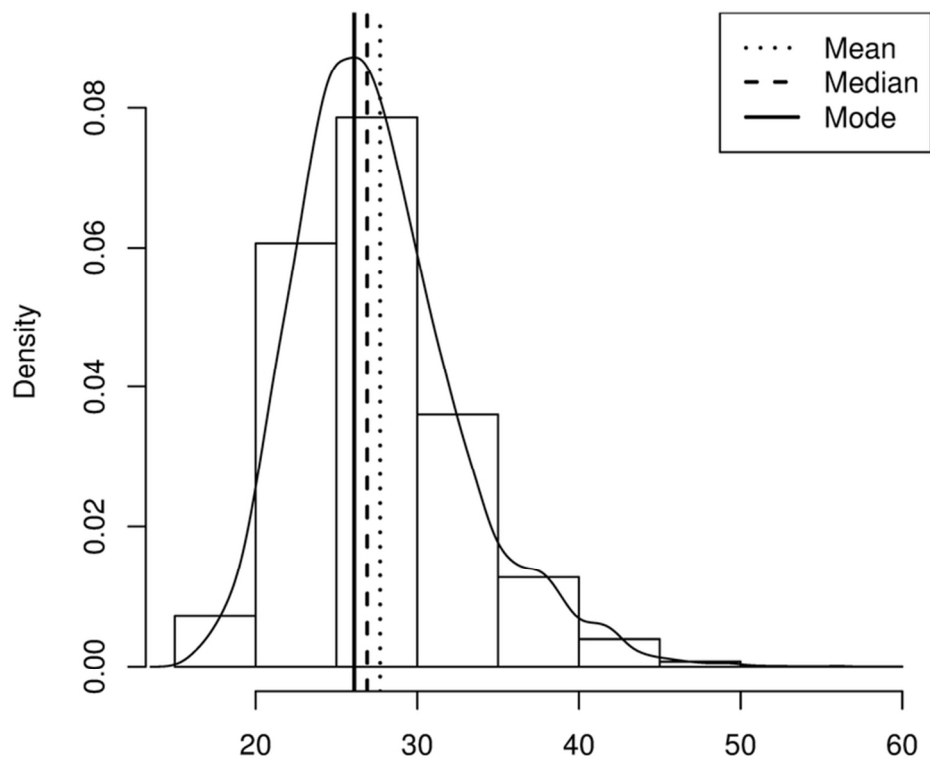


145x145mm (150 x 150 DPI)



1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

### BMI - total



145x145mm (150 x 150 DPI)

