A Perturbation Technique for Sample Moment Matching in Kernel Density Estimation

Arnab Maity¹ and Debapriya Sengupta²

Abstract

The fundamental idea of kernel smoothing technique can be recognized as one-parameter data perturbation with a smooth density. The usual kernel density estimates might not match arbitrary sample moments calculated from the unsmoothed data. A technique based on two-parameter data perturbation is developed for sample moment matching in kernel density estimation. It is shown that the moments calculated from the resulting tuned kernel density estimate can be made arbitrarily close to the raw sample moments. Moreover, the pointwise rate of MISE of the resulting density estimates remains optimal. Relevant simulation studies are carried out to demonstrate the usefulness and other features of this technique. Finally, possible extensions to estimating equation type of constraints and multivariate densities are discussed.

KEY WORDS: Kernel density estimate, Sample moment matching, Perturbation technique, Pointwise MISE

¹Department of Statistics, Texas A&M University, 3143 TAMU, College Station, TX 77843-3143.
²Theoretical Statistics and Mathematics Division, Indian Statistical Institute, 203 B. T. Road, Kolkata 700 108, INDIA
1 Introduction

In this paper we consider the problem of moment matching of the nonnegative kernel density estimate (KDE),

\[ \hat{f}_n(x) = \frac{1}{nh} \sum_{i=1}^{n} K \left( \frac{x - X_i}{h} \right), \]  

for a random sample \( X_1, X_2, \ldots, X_n \) coming from a population with density \( f \). Here \( K \) is the “kernel function”, which is taken to be a probability density function. The scaler \( h \) is “bandwidth” that controls the extent of smoothing applied to data. At the outset we assume that all moments of \( f \) are finite. There is a great deal of literature concerning the bias and variance of KDE and the strategies for choosing the bandwidth parameter \( h \). We refer to Silverman (1986), Härdle (1990), Scott (1992), Wand and Jones (1995) among others for extensive literature on this subject. In relatively small samples, the bias of KDE is an important issue as the moments calculated from \( \hat{f}_n \) can be quite different from the sample moments. The argument given by Efron and Tibshirani (1996) in favor of moment matching is “... it allows the nonparametric smoother to use a substantially greater window width without badly degrading the overall fit to the data”. The theorem 2 of this paper makes this statement more precise. Finally, from a practical point of view, if we consider density estimates as carriers of descriptive statistical information it is a good idea to pack it with more accurate higher moment information. This, for example, should make kernel density estimates more useful for survey practitioners.

Moment matching, in general, can be seen as a problem of constructing a random variable \( Y \) such that (i) the conditional distribution \( Y \) given \( X_1, X_2, \ldots, X_n \) is known and has a smooth density and, (ii) for a finite set \( S \) of positive integers and a tolerance sequence \( \tau_n \rightarrow 0 \),

\[ \left| E(Y_j | X_1, X_2, \ldots, X_n) - n^{-1} \left( \sum_{i=1}^{n} X_i^j \right) \right| \leq \tau_n \]  

for each \( j \in S \) almost surely. Of course, the desired density estimate is the conditional density of \( Y \) given \( X_1, X_2, \ldots, X_n \). In order to retain consistency and other desirable mean squared error properties \( Y \) must be obtained through small perturbations of the empirical distribution of the data. Notice that the specified moments match exactly if \( \tau_n \) can be made 0. Stated in this form, the actual problem can be easily interpreted as one closely linked with bootstrapping. The moment matching condition (2) is based on usual bootstrap asymptotics (cf. Hall, 1992), which can be altered to some other measure of probabilistic closeness if so desired.

For KDE, the basic random variable is defined by

\[ Y = X + h\epsilon \]  

(3)
where $X$ and $\epsilon$ are independent with $\epsilon$ having a density given by the kernel $K$. By defining $X^n = (X_1, X_2, \ldots, X_n)$, the bootstrap distribution of $X$ is given by $\Pr(X = X_i | X^n) = n^{-1}$ for $1 \leq i \leq n$. The random variable in (3) can be interpreted as a probabilistic data perturbation of $X^n$ with $h$ as perturbation parameter. The necessity of a smooth estimate stems from the fact that empirical data does not provide any unbiased estimate of $f$ and that the nonparametric maximum likelihood leads to an inconsistent estimate. Moments calculated under (1) can be quite different from the raw sample moment estimates provided by empirical data. For example,

$$\text{Var}(Y | X^n) = s^2_x + h^2 \sigma^2_\epsilon > s^2_x,$$

where $s^2_x$ denotes the sample variance of the raw data and $\sigma^2_\epsilon$ denotes the variance of $\epsilon$. This overestimation is significant if $h^2 \sigma^2_\epsilon$ is considered to be so. Jones (1991) investigates the problem of variance inflation and suggests two appropriate scale corrections that match the first two sample moments. One of Jones’ correction amounts to choosing

$$Y = X' + h \epsilon,$$

where $X' = \bar{X} + s^{-1}_x (s^2_x - h^2 \sigma^2_\epsilon)^{1/2}(X - \bar{X})$. This technique, however, does not work with higher moments in general. A careful inspection of possible intuition behind the above idea tells us it might be possible to match moments more accurately by considering the following two-parameter family of perturbations

$$Y = \Phi(X, \eta) + h \epsilon. \tag{4}$$

The function $\Phi$ denotes a smooth family of perturbations satisfying $\Phi(x, 0) = x$. Different techniques of reducing bias of $\hat{f}_n(x)$ are available in the literature. Among them the concept of data sharpening in Choi and Hall (1999) is the closest to our approach. However, the aspect of bias we are interested in here is quite different. Other works on bias reduction of KDE include transformed kernel estimation approach considered by Wand, Marron and Ruppert (1991) and Ruppert and Cline (1994). See also, Samiuddin and El-Sayyad (1990) and Jones, Linton and Nielsen (1995) for other approaches. Here we show that by appropriately choosing $\Phi$ in (4) as a Taylor series in $\eta$ it is possible to obtain smooth density estimates which can match a given set of raw sample moments with arbitrary precision in large samples. For smaller sample sizes we demonstrate through simulation study that the method produces reasonably accurate matching. It can also be seen that the algorithm for finding a suitable $\Phi$ is solvable by dynamic programming. It seems that the exact moment matching is theoretically possible by iterating our algorithm indefinitely. However, we do not pursue that problem in this presentation.
Hall and Presnell (1999) consider the problem of matching constraints more general than moments from a resampling perspective, as well. They look into the family of perturbations given by

\[ Y = X^\dagger + h\epsilon, \]

where \( X^\dagger \) is independent of \( \epsilon \) with \( \Pr(X^\dagger = X_i|X^n) = p_i \) for a probability vector \( p = (p_1, p_2, \ldots, p_n) \), which can be viewed as perturbation parameters. In their approach an optimal \( p \) is determined after minimizing a suitable divergence measure from (1) under given constraints. The algorithm solves the Lagrangian using Newton-Raphson iterations. It should be noted that this procedure requires solution for \( (n-1) \) unknowns (namely, \( p_1, p_2, \ldots, p_{n-1} \)) for a given data. As shown later, while dealing only with moments we can achieve very accurate matching by solving much lesser number of unknowns using our procedure.

One important criterion for such perturbation based methods is their robustness. Braun and Hall (2001) formulated the problem of robustness in this context as minimization of some distance based measure subject to the data sharpening objectives. The basic approach is bootstrap based (as in Hall and Presnell, 1999) where one needs to solve \( O(n) \) equations. In our case the moment functionals are not robust to begin with. As a consequence we observe that the maximum perturbation band increases if one goes on matching higher and higher moments. Increasing sample size reduces the maximum perturbation band.

Finally, a remark about nonnegativity of the density estimates. It can be easily seen that using higher-order kernels (Bartlett, 1963) one can match arbitrary number of moments by modifying the density estimate to

\[ g(x) = (1 + a_1 x + a_2 x^2 + \ldots) \hat{f}_n(x), \]

The coefficients \( (a_i) \) can be quickly found out by expansion through orthogonal polynomials corresponding to \( \hat{f}_n \) and the moment equations. Further, it can be shown that the coefficients tend to zero as \( h \to 0 \). Unfortunately, \( g \) may not remain nonnegative.

The paper is organized as follows. In section 2, the details of the perturbation algorithm is provided. The theoretical properties of the resulting density estimate and simulation results are described in section 3. The proofs are given in the Appendix.

2 Method

Suppose our objective is to match first \( k \) moments in (4). For the remaining of this paper we shall denote the conditional expectation given \( X_1, X_2, \ldots, X_n \) by \( \mathcal{E}^n \). Then, the moment equations are given
by
\[ \mathcal{E}^n (\Phi(X, \eta) + h \epsilon)^j = n^{-1} \sum_{i=1}^{n} X_i^j, \quad (5) \]
for \( 0 \leq j \leq k \). For the sake of operational convenience we choose \( \eta = h \) from now on. However, this is not necessary. Using Binomial expansion we can rewrite the system of equations given by (5) in matrix notation as
\[ H^{-1} AZ = \beta_n, \]
where \( H = \text{diag}(1, h, \ldots, h^k) \), \( Z = (1, E^n \Phi, \ldots, E^n \Phi^k)' \), and \( A = ((a_{jr})) \) is a \((k+1) \times (k+1)\) matrix such that, \( a_{00} = 1; a_{jr} = \binom{j}{r} \epsilon^j r \) for \( j \geq 1 \) and \( 0 \leq r \leq j \); \( a_{jr} = 0 \) if \( r > j \). Next, by solving \( z \) in the above equation we get the following system of equations which we primarily use for constructing the perturbation function \( \Phi \).
\[ \mathcal{E}^n \Phi^j(X, h) = \sum_{r=0}^{j} c_{n,jr} h^r, \quad 0 \leq j \leq k, \quad (6) \]
where \( (c_{n,jr}) \) are known constants depending on first \( k \) sample moments, the moments of \( \epsilon \) and \( h \) (for \( j = 0 \) the equation is trivial). Note also that \( c_{n,jr} = 0 \) for \( r > j \). The structure of the system of equations given by (6) intuitively suggests the following nature of the solution.
\[ \Phi(X, h) = \sum_{r=0}^{T} h^r \phi_r(X) \quad \text{where} \quad \phi_0(X) = X. \quad (7) \]
The Taylor expansion (about \( h = 0 \)) in (7) allows one to equate the coefficients of \( h^l \) on two sides of (6) and consequently, solve the coefficient functions \( \phi_r(X) \) unknown so far. Although the above method does not guarantee an exact solution it provides a reasonably accurate solution for each \( h > 0 \). The nonlinear nature of (6) is the root cause behind the difficulty in deriving exact solutions. In what follows, we show that by taking \( T \) large the error of the approximate solution can be made arbitrarily small. However, increasing \( T \) also adversely affects the performance of the algorithm.

Next define
\[ S_r(j) = \left[ \frac{d}{dh} \right]^r \Phi^j(X, h) \bigg|_{h=0}. \]
Using (6) and (7) we can write
\[ \mathcal{E}^n \Phi^j(X, h) = \sum_{r=0}^{jT} (j!)^{-1} \mathcal{E}^n (S_r(j)) h^r = \sum_{r=0}^{j} c_{n,jr} h^r, \quad (8) \]
by (6).

Note that \( S_r(0) \) is equal to zero for each \( r \geq 1 \) and \( S_0(j) = X^j \) for each \( j \geq 1 \). A direct computation of \( S_r(j) \) is tedious and computationally inefficient. However, an efficient management of these data
objects is very crucial in solving (6). In the following lemma we derive a recursion formula for \( S_r(j) \) that helps our cause in two ways. Firstly, it shows that computation of \( S_r(j) \) can be done using dynamic programming. Secondly, the very nature of the recursion formula suggests a derived set of equations for the coefficient functions \( \phi_r(X) \).

**Lemma 1.** For \( j \geq 1 \), the quantities \( S_r(j) \) defined above satisfies the following.

(i) For \( r = 1 \), \( S_1(j) = j \phi_1(X) X^{j-1} \).

(ii) For \( r \geq 2 \),

\[
S_r(j) = j \alpha! \phi_\alpha(X) X^{j-1} + j \sum_{l=1}^{a-1} \frac{(r-1)!}{(r-l)!} \phi_l(X) S_{r-1}(j-l).
\]

where \( \alpha = \min(T, r) \).

Technical details are provided in the appendix. Using (6), (8) and (9), we now describe the main steps of the *perturbation algorithm* by equating coefficients of \( h^r(1 \leq r \leq T) \), in (8).

1. Obtain \( \phi^*_n(X) \) as a \((k-1)\)-degree polynomial in \( X \) satisfying

\[
\mathcal{E}^n \phi^*_n(X) X^{j-1} = (j-1)! c_{n,j}
\]

for \( 1 \leq j \leq k \). Hence, obtain corresponding \( S^*_n(j), 0 \leq j \leq k \) from lemma 1(i).

2. After obtaining \( \{ \phi^*_nm, S^*_nm(0 : 1 : k), 1 \leq m \leq r-1(r \geq 2) \} \), determine \( \phi^*nr(X) \) as a \((k-1)\)-degree polynomial in \( X \) by solving

\[
\mathcal{E}^n \left( \phi^*nr(X) X^{j-1}(X) \right) \frac{1}{\mathcal{P}n(j)} \left\{ c_{n,jr} - j \sum_{l=1}^{r-1} \frac{(r-1)!}{(r-l)!} \mathcal{E}^n \left( \phi^*_nl(X) S^*_nm(r-l)(j-l-1) \right) \right\},
\]

for \( 1 \leq j \leq k \). Hence, obtain \( S^*_nr(0 : 1 : k) \).

3. Iterate until \( r = T \).

The required perturbed kernel density estimate will then be given by the conditional density of

\[
Y = \Phi^*_n(X, h) + h \epsilon = X + \sum_{r=1}^{T} h^r \phi^*_nr(X) + h \epsilon
\]

given \( X_1, X_2, \ldots, X_n \). Each \( \phi^*_r(X) \) is a \((k-1)\) degree polynomial in \( X \) whose coefficients have been computed by the above steps. The estimate can be rewritten as

\[
\hat{f}^*_n(x) = (nh)^{-1} \sum_{i=1}^{n} K \left( \frac{(x - \Phi^*_n(X_i, h))}{h} \right).
\]
3 Results

In this section we discuss some theoretical properties of the perturbation algorithm for moment matching. Specifically, we focus on (i) bias properties of the perturbed KDE in context of (2), (ii) large sample rate of the pointwise MSE of (12) and, (iii) Simulation Studies.

3.1 Moment Matching Bias

The following proposition summarizes some useful properties of the solution of the perturbation algorithm given in (11).

Proposition 1. (i) One can find finitely many sample averages \(G_n = (\bar{g}_n^1, \bar{g}_n^2, \ldots, \bar{g}_n^b)\) → \(\gamma\) almost surely, such that for each \(1 \leq r \leq T\), \(\phi^*_nr(x) = \sum_{l=0}^{k-1} W_{rl}(G_n) x^l\) for suitable functions \(W_{rl}\), which are continuously differentiable in a neighborhood of \(\gamma\).

(ii) For each \(0 \leq j \leq k\) and \(1 \leq r \leq jT\), \(\mathcal{E}^n S^*_n(r)\) (and consequently, \(\mathcal{E}^n \Phi^*_n(X, h)\)) can be expressed as a function of \(G_n\) which is also continuously differentiable in a neighborhood of \(\gamma\). Moreover, \(\mathcal{E}^n S^*_n(r) = j! c_{n,jr}\) for \(1 \leq r \leq T\).

(iii) If the kernel \(K\) is symmetric then \(\phi^*_n(2l+1) \equiv 0\) for \(1 \leq l \leq [(T-1)/2]\).

The proof is given in the appendix. The proposition is useful in making a few important observations. Firstly, from (i) above we can see that \(\Phi^*_n\) obtained from the perturbation algorithm has a population analogue by the strong law, namely,

\[
\Phi^*_n(x, h) \overset{a.s.}{\rightarrow} x + \sum_{r=1}^{T} \left( \sum_{l=0}^{k-1} W_{rl}(\gamma) x^l \right) h^r
\]

\[
= x + \sum_{l=0}^{k-1} \left( \sum_{r=1}^{T} W_{rl}(\gamma) h^r \right) x^l
\]

(13)

for every \(h > 0\). Thus, the perturbation technique can be thought of as an estimation procedure. From the proof of proposition 1 it may be observed that the choice of \(\phi^*_nr\) in (10) is unique in the space of polynomials of degree \(\leq (k-1)\). The perturbation technique actually has a close similarity with the idea of projection as given by Hall and Presnell (1999). To see this, think of the space \(\mathcal{V}_k\) of densities whose moments up to order \(k\) match with that of \(f\). The expected value of KDE is given by the convolution \(f_n = f \ast \{h^{-1}K(\cdot/h)\}\), which does not belong to \(\mathcal{V}_k\). Following Hall and Presnell (1999), the natural thing to do in this context would be project \(f_n\) onto \(\mathcal{V}_k\). The perturbation family of densities, generated
by (4), provides parametric approximations of the ideal projection without deviating significantly from $f_n$. This establishes a philosophical connection between our approach and that of Efron and Tibshirani (1996), as well. Any natural distance function or divergence measure that is being optimized in our technique is not clear at this point. However, the equations in (10) closely resemble normal equations in least squares.

The part (iii) of proposition 1 implies that the usual bias properties (namely, $O(h^4)$) of the perturbed density estimate in (12) is not seriously altered due to moment matching (to be clarified later). Because the algorithm provides moment matching for each $h > 0$ the implementation of the algorithm with optimal bandwidth for the usual KDE is obvious. Before going into the further details regarding pointwise bias of perturbed KDE we turn to the problem of moment matching. The mathematical formulation of the general problem has already been presented in section 1. The following theorem describes the extent of moment matching of tuned kernel density estimates.

**Theorem 1.** Let $X^n = (X_1, X_2, \ldots, X_n)$ be a random sample of size $n$ from a univariate distribution with density $f$. Assume that all moments of $f$ are finite. Consider the perturbed KDE given by (12) with bandwidth $h_n$. Then, for a bandwidth sequence $h_n \to 0$ we have,

$$\left| E(Y^j|X^n) - n^{-1} \left( \sum_{i=1}^{n} X_i^j \right) \right| = o(h^T_n)$$

almost surely as $n \to \infty$, for $1 \leq j \leq k$.

In view of the above theorem it is natural to ask what happens if one exactly matches the moments by taking $T \to \infty$. Intuitively, such a result should be true at least for densities with compact support. A rigorous proof seems to be difficult.

Next, we look into the issue of MISE of (12). The important conclusion is that the optimal rate of convergence remains unaltered by the proposed perturbation scheme. The sketch of the proof is given in the appendix.

**Theorem 2.** If the underlying density $f$ is twice continuously differentiable and if $h_n \approx C_0 n^{-1/5}$, then $E \left( \tilde{f}_n(x) - f(x) \right)^2 = O(n^{-4/5})$.

### 3.2 Simulations

In this section, we study various performance characteristics of the simple KDE and the one obtained through perturbation technique using simulation. Following four distributions are considered.

2. Symmetric bimodal normal mixture : \( \frac{1}{2}N(-1, 4/9) + \frac{1}{2}N(1, 4/9) \).

3. Asymmetric bimodal normal mixture : \( \frac{3}{4}N(0, 1) + \frac{1}{4}N(3/2, 1/9) \).

4. Chi-squared with 6 degrees of freedom.

For each distribution, samples of size 50, 100 and 200 are taken respectively. Density estimation is done by simple KDE (zeroth moment matched) and matching up to fourth and sixth moments as well. The Epanechnikov kernel (Silverman, 1986) is used. The optimal bandwidth is taken using the formula

\[
h_n^* = \left( \frac{\int (K(t))^2 dt}{\int (f^{(2)}(x))^2 dx \int t^2 K(t) dt^2} \right)^{1/5} n^{-1/5}
\]

The whole process is simulated 1000 times and median of pointwise estimates are recorded. In each case, average absolute moment biases (for 2nd, 3rd, 4th moments), average absolute percentile biases (for 10th, 25th, median, 75th, 90th percentiles) are given in tabular form. Also, the average Hellinger distance and MISE are calculated. The Hellinger distance is chosen in order to over-penalize the errors in the tails. For each of the sample sizes considered, plots of density estimates are given. Also for each distribution, plot of perturbation transformation is given for matching up to 4th and 6th moments.
3.2.1 Standard Normal

In the case of standard normal distribution, moment matched estimates are better than the simple KDE both in the terms of moment biases and Hellinger’s distance. Different statistics are given in Table 1. It may be noted that moment biases are significantly less if we match up to fourth moment. But when up to sixth moment is matched, bias in the fourth moment increases compared to the case when only up to fourth moment is matched. Hellinger distances are more or less same in all three cases (for each sample size). From the percentile bias section of Table 1 we see that percentile biases are reduced in all cases. A visual perception can be obtained from Figures 1(a)-(c). A plot of perturbation function, $\Phi$, is given in Figure 5(a). From Figure 5(a), it can be seen that, matching up to sixth moment yields more perturbed data towards two tails than in case of matching up to fourth moment.
3.2.2 Symmetric Bimodal Normal Mixture

This density is underestimated by both simple KDE and our technique near the local minima at the median. Performance of both the techniques in case of this distribution are given in Table 2. From the moment bias section of Table 2, it can be seen that moment biases are becoming significantly less compared to bias generated by simple KDE. But here fourth moment bias does not increase when up to sixth moment are matched (which was the case in previous section). A visual perception can be obtained from Figures 2(a)-(c). Plot of perturbation function is given in Figure 5(b). From the plot it can be seen that, when matched up to fourth moment data is less perturbed towards the two tails. But when matched up to sixth moment, perturbation is more towards the tail region.

3.2.3 Asymmetric Normal Mixture

While estimating asymmetric normal mixture, both methods estimate left tail quite satisfactorily but do not give good estimate at right tail. Performance of both the methods are given in Table 3. From the moment bias section of Table 3, it can be seen that moment biases are decreasing when up to fourth or sixth moments are matched. For each sample size, biases decrease as number of moment matched is increased from four to six. If Hellinger distances are considered, it can be seen that they are larger compared to other cases. From the percentile bias section of Table 3, it can be seen that biases are reducing here also. Unlike other cases, here median bias does not change significantly from other percentile biases. But 90th percentile bias is changing significantly when moment matching method is applied. Plots for density estimates for various sample sizes 50, 100, 200 are given in Figure 3(a)-(c). Plot of perturbation function is given in Figure 5(c). Like in the case of symmetric normal mixture, when matched up to fourth moment, amount of perturbation towards the tails becomes small compared to the case when matched up to sixth moments.

3.2.4 Chi-squared(with 6 degrees of freedom)

Different statistics are given in Table 4. Here relative moment bias is supplied instead of actual moment bias. Also, while matching up to sixth moment number of polynomials($\phi_i$) was increased to 8 i.e. considering an error of $o(h^8)$. Here also moment biases are reducing compared to that of simple KDE. In general it can be said that moment biases are decreasing when number of moments matched is increased from four to six(except the case, where sample size = 50 and second moment bias is considered). From the Hellinger distance column we see that when matched up to fourth moments, this yields least Hellinger distance among all four distributions considered(in respective cases). Looking at the percentile
bias section of Table 4 we can see that percentile bias remains consistent in most of the cases except in case of 25th percentile. Plots for density estimates for various sample sizes are given in Figure 4(a)-(c). It can be seen that, when matched up to sixth moment the density estimate is deformed near the peak region of the density though moment bias is decreasing. But this feature is reflected in the percentile bias section of Table 4. It can be seen that, when matched up to sixth moment, 25th percentile bias is greatest for each sample size considered. But as sample size increases the "deformity" decreases. It is evident from the plots and as well as from the table also as percentile biases are gradually decreasing.

Plot of perturbation function is given in Figure 5(d). It can be seen that data is more perturbed towards the tails. Moreover, matching up to sixth moment yields more perturbed data compared to data given by matching up to 4th moment. Also, perturbation function lies above the $y = x$ line at the left tail whereas it lies below (much below in case of matching up to sixth moment) towards the right tail.

### 3.2.5 Conclusions and Other Issues

From estimates and different statistics of the distributions considered above, we see that, moment biases are becoming less compared to the bias generated by doing simple kernel density estimation. It can be noted that for each moment, moment bias is decreasing as number of moment matched is increased from four to six. Though moment bias is decreasing when matched up to 6th moment, a slight increment in Hellinger’s distance can be noted in most of the cases for all sample sizes considered.
Looking at percentile biases, it can be seen that percentile biases are lower than that of simple kernel density estimates in most of the cases. When matched up to 6th moment, heavy tailed distributions like Chi-squared gives a “deformed” estimate which can be thought as a force driving the estimate to some sort of symmetry. This deformity becomes lower as sample size increases. But matching up to fourth moment gives reasonably good estimate. These feature is reflected clearly when Hellinger’s distance is taken into account. MISE is less sensitive in this case.

Next, we discuss certain robustness issues related to our procedure. Since the perturbation function $\Phi$ in (7) is approximated by polynomials it is expected that the extent to which the data points shift is related to the degree of the polynomials in $\phi_r$. When we match higher moments the degree of these elemental polynomials also increase and the procedure starts losing its robustness. This is clearly visible from Figure 4 where the procedure breaks down when six moments are matched. The reality is however more complex. With increasing sample sizes optimal bandwidths decrease and with smaller bandwidths amount of perturbation is less (thus, the procedure is more stable). It might be possible to introduce certain penalties for large perturbations. If we compare this procedure with Braun and Hall (2001) (see also, Hall and Presnell, 1999) we find that the issue of breakdown in their procedure is dependent on assigning undue bootstrap weights ($p_i$) to extreme observations. This guards against shifting the data to outside the sample range for heavy tailed distributions. In this respect their procedure seems to be
more robust than ours for higher moment matching (more than six) for relatively smaller sample sizes. It is noteworthy that the asymptotic formulas are unable to explain this situation.

In view of our simulations (Figure 6) it seems that overall, it is safer to match up to four moments. If the underlying density is not heavy tailed one may venture with six moments (Figure 6(a)-(i)). However, if the density is heavy tailed the average amount of perturbation becomes higher rendering the estimate unrealistic. In Figure 6, we consider maximal absolute perturbations in order to study sensitivity towards large displacements of the original data (even if they occur with low probability). If one considers an average measure of displacements (like, Braun and Hall, 2001) it will be less sensitive towards really large displacements.

Finally, we compare our method to the perturbation methods proposed by Jones (1991). Two perturbation formulae were proposed,

Method I. Replace the data, \( X \), by \( X' = \bar{X} + (s^2 + h^2)^{-1/2}s(X_i - \bar{X}) \) and employ the new bandwidth \( h_v = s(s^2 + h^2)^{-1/2}h \) where \( s \) is sample standard deviation.

Method II. Replace \( X \) by \( W = \bar{X} + s^{-1}(s^2 - h^2)^{1/2}(X_i - \bar{X}) \) with the same bandwidth.

We calculate MISE and Hellinger distance of the estimates for each of the four distributions considered in this section when 2nd moment is matched using our method and compare these values with those produced by Jones’ methods. Results for comparison with method-I are given in Table 5. Comparison results for method-II yeilds same MISE and Hellinger’s distance for both the procedures as the procedure proposed in this paper is same as method-II when matched up to 2nd moment. In view of our simulation results from Table 5 it is clear that our method proposed in this paper has an advantage over method-I in terms of improved MISE and Hellinger distance.
Figure 6: Boxplots of distributions of maximum absolute perturbations over 1000 simulations for fourth and sixth moments matching with sample sizes 50, 100 and 200 respectively. 6(a)-(c) standard normal distribution; 6(d)-(f) symmetric normal mixture; 6(g)-(i) asymmetric normal mixture and 6(j)-(l) chi-squared distribution (df = 6).
References


4 Appendix

A.1. Proof of Lemma 1.

It is easy to verify (i). For (ii)
\[ S_r(j) = \left[ \left( \frac{d}{dh} \right)^r \Phi^j(h) \right]_{h=0} = \left[ D_r \Phi^j(h) \right]_{h=0} = \left[ D_{r-1} \left( j \Phi^{j-1}(X,h), \frac{d}{dh} \Phi(X,h) \right) \right]_{h=0} \]
\[ = \left[ D_{r-1} \left( j \Phi^{j-1}(X,h) \sum_{l=1}^{T} l \phi_l(X) h^{l-1} \right) \right]_{h=0} = \left[ j \sum_{l=1}^{T} l \phi_l(X) D_{r-1} \left( h^{l-1} \Phi^{j-1}(X,h) \right) \right]_{h=0} \]
\[ = j \sum_{l=1}^{\min(T,r)} l \phi_l(X) \left( \frac{r-1}{l-1} \right) (l-1)! S_{r-l}(j-1) \]
\[ = r! j \phi_r(X) X^{j-1} + j \sum_{l=1}^{r-1} \frac{(r-1)!}{(r-l)!} \phi_l(X) S_{r-l}(j-1) \]

The last line of expressions is valid when \( r \leq T \).


We begin by noting some basic facts. Firstly, the quantities \( c_{n,jr} \) are linear (hence continuous) functions of the sample moments \( \beta_n \) in view of (5) and (6). Next, if we plug in \( \phi_{nr}(x) = \sum_{k=0}^{k-1} f_l x^l \) in (10), the solution for the coefficient vector \( \beta_n \) in (10) is given by \( \gamma^{n}X^{p+q-1} \). Because the observations are coming from a density, \( EV_n (= EV_1) \) is invertible and by strong law, \( V_n \rightarrow EV_1 \) almost surely. Next, by the inverse function theorem, analyticity of the map \( A \rightarrow \det(A) \) and Cramer’s rule formula for computing \( A^{-1} \) one can see that the map \( A \rightarrow A^{-1} \) is continuously differentiable function of the arguments of \( A \) in a neighborhood of \( EV_1 \). Thus, the entries of \( V_n^{-1} \) can be expressed as a continuously differentiable function of the entries of \( V_n \) (which are sample averages) almost surely eventually. In view of these we prove (i) if we can show that \( d \) above can be expressed as a smooth function of finitely many sample averages. We sketch an induction argument for that. For \( r = 1 \), it is easy to see from step 1 of the perturbation algorithm that the claim is true. Also, note that the same claim will be true for \( S_{r+1}(j) \) in view of lemma 1. If the claim is assumed to be true for \( \phi_{nm}^*, S_{nm}^*(0 : 1 : k), 1 \leq m \leq (r-1) \), then the claim should also be true for \( r \). To see this, firstly, the argument related to \( V_n \) continues to apply; secondly, the right hand side of (10) is clearly a linear combination (hence smooth) of a (finite) new set of sample averages. Hence (i) is proved. (ii) is a direct conclusion of lemma 1, (8) and also above arguments.

Proof of (iii) is again by induction. First we prove \( \phi_{i}^* \equiv 0 \) \( (l = 0) \). Elementary arguments show that \( c_{j1} = 0 \) for \( 1 \leq j \leq k \) using the symmetry of the kernel \( K \). Therefore,
\[ EX^{j-1} \phi_{r}^*(X) = 0 \quad \text{for} \quad j = 1, 2, \ldots, k. \]

Hence, \( \phi_{r}^*(X) \equiv 0 \). Our claim is true for \( l = 0 \). Let us assume that the claim is true for \( l = 0, 1, \ldots, m-1 \).
Want to prove for $l = m$. To calculate $\phi_{2m+1}^*$ compare the coefficients of $h^{2m+1}$ in (9) and get

$$E \sum_{A} \prod_{i=0}^{n} \phi_{i}^{* \alpha_{i}} (X) = 0 \quad (\phi_{0}^{*} (X) = X)$$

where the summation is taken over $(T+1)$-tuples nonnegative integers, $A = \{ (\alpha_{0}, \ldots, \alpha_{T}) \geq 0 : \alpha_{0} + \alpha_{1} + \ldots + \alpha_{T} = j, \alpha_{1} + 2\alpha_{2} + \ldots + Ta_{T} = 2l + 1 \}$. The left hand side is obtained from (8) by expanding $\Phi^{*j}$ algebraically. A term in the above sum is $EX^{j-1} \phi_{2m+1} (X)$ as

$$\alpha_{0} = j - 1, \alpha_{1} = \ldots = \alpha_{2l} = 0, \alpha_{2l+1} = 1, \alpha_{2l+2} = \ldots = \alpha_{n} = 0$$

belong to $A$. Now we observe that other terms in the sum are themselves zero, because there exists at least one such $\alpha_{i}$ such that $i = 2k + 1$ for some $k \leq m - 1$ as $2l + 1$ is odd. Hence every term has a multiplier of the form $\phi_{2k+1}^{*} (X)$ which is zero by our assumption. So, we get

$$EX^{j-1} \phi_{2m+1} (X) = 0 \quad \text{for} \quad j = 1, 2, \ldots, k$$

Hence we get $\phi_{2m+1}^{*} = 0$. This completes the proof.


For $k = 1$ the result is obvious as $E\phi_{n1}^{*} (X) \equiv 0$. Assume $k \geq 2$. In view of coefficient matching in (9), we get

$${\mathcal E}^{n} \Phi_{n}^{* j} = \sum_{l=1}^{T} c_{n,j}h_{n}^{l} + h_{n}^{T} \sum_{l=T+1}^{jT} d_{n,j}h_{n}^{(l-T)}, \quad (15)$$

where $d_{n,j}$ are smooth functions of the moment vector $\beta_{n}$ in view of fact 1. By an application of strong law of large numbers we get that the second sum in (15) is $o(h_{n}^{-2})$ almost surely (as we are dealing with only finite number of terms). Let $z_{n} = (1, \mathcal{E}^{n} \Phi_{n}^{* 1}, \ldots, \mathcal{E}^{n} \Phi_{n}^{* k})'$. Therefore, using (7) we can write

$$z_{n} = H_{n}^{-1} A^{-1} H_{n} \beta_{n} + o(h_{n}^{-2})$$

almost surely. Here, $H_{n} = \text{diag} (1, h_{n}, \ldots, h_{n}^{k})$ and $A$ as defined in (7). Because $A$ is a lower triangular by definition, $H_{n} A^{-1} H_{n}^{-1}$ is a continuous function (hence bounded as $h_{n} \to 0$). Therefore,

$${\mathcal E}^{n} (\Phi_{n}^{*} (X, h_{n}) + h_{n}c)^{j} = H_{n} A H_{n}^{-1} z_{n} = \beta_{n} + o(h_{n}^{-2})$$

almost surely. Hence the theorem.


We shall skip the rigorous details here and present only heuristics. Technical assumptions will be mentioned as we go along. In what follows the gaps can be filled using standard techniques (see, H"ardle, 1990 for example). First, note that

$$E \left( \hat{f}_{n}^{*} (x) - f (x) \right)^{2} \leq 3 \left[ E \left( \hat{f}_{n}^{*} (x) - f_{n}^{*} (x) \right)^{2} + E \left( f_{n}^{*} (x) - \hat{f}_{n} (x) \right)^{2} + E \left( \hat{f}_{n} (x) - f (x) \right)^{2} \right] \quad (A1)$$

18
where \( f_n^*(x) = (nh)^{-1} \sum_{i=1}^{n} K ((x - \Phi^*(X_i, h))/h) \) with \( \Phi^*(x, h) \) defined by (13) and \( \hat{f}_n(x) \) as in (1).

Our job is to get estimates of the first two terms on the right hand side of (A1) as third one is the standard term. Now, assume that the kernel \( K \) is symmetric and has a bounded, continuous first derivative. By proposition 1,

\[
\Phi_n^*(x, h) = x + h^2 \phi_{n2}(x, h) + \ldots + h^{2\lfloor T/2 \rfloor} \phi_{n(2\lfloor T/2 \rfloor)}(x, h). \tag{A2}
\]

We apply a very crude bound on the first term using Cauchy-Schwarz inequality

\[
E \left( \hat{f}_n^*(x) - f_n^*(x) \right)^2 \leq n^{-1} h^{-2} \sum_{i=1}^{n} E \left\{ K \left( (x - \Phi_n^*(X_i, h))/h \right) - K \left( (x - \Phi^*(X_i, h))/h \right) \right\}^2
\]

\[
\leq C_0 n^{-1} h^{-4} \sum_{i=1}^{n} E \left\{ \sum_{2 \leq 2r \leq T} \left( \sum_{l=0}^{k-1} (W_{2r+l}(G_n) - W_{2r+l}(\gamma)) X_l \right) \right\}^2 h^{2r}
\]

\[
\text{(from (14), (A2) and the mean value theorem on } K)
\]

\[
\leq C_0 n^{-1}.
\]

The last step follows from the routine arguments used for delta methods (cf. Serfling, 1980, pp.122-123) in view of proposition 1.

For the second term let \( v(x) = -(\Phi^*(x) - x)/h^2 \) and \( \xi_i = K ((x - X_i)/h + hv(X_i)) - K ((x - X_i)/h) \).

Thus, \( \xi_i \)'s are iid and

\[
E \left( f_n^*(x) - \hat{f}_n(x) \right)^2 = n^{-2} h^{-2} E \left\{ \sum_{i=1}^{n} \xi_i \right\}^2 \leq h^{-2} \left( n^{-1} E \xi_1^2 + (E \xi_1)^2 \right).
\]

It is not difficult to see that \( E \xi_1^2 = O(h^3) \) and \( E \xi_1 = O(h^3) \). Therefore, \( E \left( f_n^*(x) - \hat{f}_n(x) \right)^2 \leq C_1 n^{-1} h + C_2 h^4 \) for suitable constants \( C_1 \) and \( C_2 \). The third term in (A1) is a standard one for kernel density estimation and it is well known (see, Härdle, 1990, for example) that with optimal choice of \( h_n \approx C_0 n^{-1/5} \), \( E \left( \hat{f}_n(x) - f(x) \right)^2 = O(n^{-4/5}) \).
Table 1: Performance of perturbed KDE technique for Standard Normal Distribution

<table>
<thead>
<tr>
<th>sample size(n)</th>
<th>matched</th>
<th>MISE</th>
<th>Hellinger distance</th>
<th>moment bias</th>
<th>percentile bias</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2nd</td>
<td>3rd</td>
</tr>
<tr>
<td>0(1)</td>
<td>0.0082</td>
<td>0.2392</td>
<td>0.1389</td>
<td>0.2116</td>
<td>0.9295</td>
</tr>
<tr>
<td>50</td>
<td>0.0070</td>
<td>0.2258</td>
<td>0.0964</td>
<td>0.0202</td>
<td>0.1611</td>
</tr>
<tr>
<td>6</td>
<td>0.0077</td>
<td>0.2464</td>
<td>0.0612</td>
<td>0.0137</td>
<td>0.1988</td>
</tr>
<tr>
<td>0(1)</td>
<td>0.0053</td>
<td>0.1868</td>
<td>0.1278</td>
<td>0.1576</td>
<td>0.8327</td>
</tr>
<tr>
<td>100</td>
<td>0.0045</td>
<td>0.1753</td>
<td>0.0763</td>
<td>0.0114</td>
<td>0.1823</td>
</tr>
<tr>
<td>6</td>
<td>0.0047</td>
<td>0.1886</td>
<td>0.0602</td>
<td>0.0092</td>
<td>0.1901</td>
</tr>
<tr>
<td>0(1)</td>
<td>0.0031</td>
<td>0.1454</td>
<td>0.1276</td>
<td>0.1079</td>
<td>0.8091</td>
</tr>
<tr>
<td>200</td>
<td>0.0026</td>
<td>0.1340</td>
<td>0.0699</td>
<td>0.0073</td>
<td>0.1825</td>
</tr>
<tr>
<td>6</td>
<td>0.0028</td>
<td>0.1441</td>
<td>0.0612</td>
<td>0.0068</td>
<td>0.1923</td>
</tr>
</tbody>
</table>

Table 2: Performance of perturbed KDE technique for Symmetric Bimodal Normal Mixture

<table>
<thead>
<tr>
<th>sample size(n)</th>
<th>matched</th>
<th>MISE</th>
<th>Hellinger distance</th>
<th>moment bias</th>
<th>percentile bias</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2nd</td>
<td>3rd</td>
</tr>
<tr>
<td>0(1)</td>
<td>0.0316</td>
<td>0.4450</td>
<td>0.1991</td>
<td>0.3232</td>
<td>1.7219</td>
</tr>
<tr>
<td>50</td>
<td>0.0341</td>
<td>0.4492</td>
<td>0.0900</td>
<td>0.0249</td>
<td>0.4198</td>
</tr>
<tr>
<td>6</td>
<td>0.0354</td>
<td>0.4613</td>
<td>0.0894</td>
<td>0.0279</td>
<td>0.4160</td>
</tr>
<tr>
<td>0(1)</td>
<td>0.0281</td>
<td>0.4158</td>
<td>0.1942</td>
<td>0.2337</td>
<td>1.7581</td>
</tr>
<tr>
<td>100</td>
<td>0.0299</td>
<td>0.4163</td>
<td>0.0911</td>
<td>0.0189</td>
<td>0.4256</td>
</tr>
<tr>
<td>6</td>
<td>0.0301</td>
<td>0.4216</td>
<td>0.0903</td>
<td>0.0174</td>
<td>0.4177</td>
</tr>
<tr>
<td>0(1)</td>
<td>0.0267</td>
<td>0.3975</td>
<td>0.1987</td>
<td>0.1716</td>
<td>1.7560</td>
</tr>
<tr>
<td>200</td>
<td>0.0275</td>
<td>0.3964</td>
<td>0.0910</td>
<td>0.0117</td>
<td>0.4215</td>
</tr>
<tr>
<td>6</td>
<td>0.0277</td>
<td>0.3991</td>
<td>0.0905</td>
<td>0.0151</td>
<td>0.4185</td>
</tr>
</tbody>
</table>
Table 3: Performance of perturbed KDE technique for Asymmetric Bimodal Normal Mixture

<table>
<thead>
<tr>
<th>sample size (n)</th>
<th>moment matched</th>
<th>MISE</th>
<th>Hellinger distance</th>
<th>moment bias 2nd</th>
<th>3rd</th>
<th>4th</th>
<th>percentile bias 10th</th>
<th>25th</th>
<th>median</th>
<th>75th</th>
<th>90th</th>
</tr>
</thead>
<tbody>
<tr>
<td>0(1)</td>
<td>0.0313</td>
<td>0.4301</td>
<td>0.1559</td>
<td>0.2008</td>
<td>1.3169</td>
<td>0.1044</td>
<td>0.0675</td>
<td>0.0610</td>
<td>0.0563</td>
<td>0.1306</td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>0.0311</td>
<td>0.4235</td>
<td>0.0783</td>
<td>0.0833</td>
<td>0.4829</td>
<td>0.0653</td>
<td>0.0437</td>
<td>0.0189</td>
<td>0.0477</td>
<td>0.0578</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>0.0324</td>
<td>0.4374</td>
<td>0.0736</td>
<td>0.0555</td>
<td>0.3161</td>
<td>0.0536</td>
<td>0.0653</td>
<td>0.0322</td>
<td>0.0360</td>
<td>0.0166</td>
<td></td>
</tr>
<tr>
<td>0(1)</td>
<td>0.0263</td>
<td>0.3802</td>
<td>0.1589</td>
<td>0.2413</td>
<td>1.3075</td>
<td>0.1246</td>
<td>0.0869</td>
<td>0.0847</td>
<td>0.0767</td>
<td>0.1417</td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>0.0267</td>
<td>0.3792</td>
<td>0.0811</td>
<td>0.0904</td>
<td>0.4816</td>
<td>0.0711</td>
<td>0.0484</td>
<td>0.0254</td>
<td>0.0527</td>
<td>0.0610</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>0.0261</td>
<td>0.3813</td>
<td>0.0754</td>
<td>0.0581</td>
<td>0.3282</td>
<td>0.0626</td>
<td>0.0596</td>
<td>0.0264</td>
<td>0.0370</td>
<td>0.0196</td>
<td></td>
</tr>
<tr>
<td>0(1)</td>
<td>0.0237</td>
<td>0.3502</td>
<td>0.1987</td>
<td>0.1716</td>
<td>1.7560</td>
<td>0.1272</td>
<td>0.0643</td>
<td>0.0790</td>
<td>0.0647</td>
<td>0.1310</td>
<td></td>
</tr>
<tr>
<td>200</td>
<td>0.0234</td>
<td>0.3461</td>
<td>0.0910</td>
<td>0.0117</td>
<td>0.1604</td>
<td>0.0540</td>
<td>0.0420</td>
<td>0.0030</td>
<td>0.0424</td>
<td>0.0545</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>0.0234</td>
<td>0.3478</td>
<td>0.0751</td>
<td>0.0572</td>
<td>0.3232</td>
<td>0.0643</td>
<td>0.0539</td>
<td>0.0217</td>
<td>0.0404</td>
<td>0.0248</td>
<td></td>
</tr>
</tbody>
</table>

Table 4: Performance of perturbed KDE technique for Chi-squared Distribution (df = 6)

<table>
<thead>
<tr>
<th>sample size (n)</th>
<th>moment matched</th>
<th>MISE</th>
<th>Hellinger distance</th>
<th>moment bias 2nd</th>
<th>3rd</th>
<th>4th</th>
<th>percentile bias 10th</th>
<th>25th</th>
<th>median</th>
<th>75th</th>
<th>90th</th>
</tr>
</thead>
<tbody>
<tr>
<td>0(1)</td>
<td>0.0041</td>
<td>0.3506</td>
<td>0.0514</td>
<td>0.0876</td>
<td>0.1385</td>
<td>0.4702</td>
<td>0.3527</td>
<td>0.3249</td>
<td>0.4104</td>
<td>0.5190</td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>0.0044</td>
<td>0.3538</td>
<td>0.0164</td>
<td>0.0199</td>
<td>0.0225</td>
<td>0.4195</td>
<td>0.0918</td>
<td>0.3076</td>
<td>0.1668</td>
<td>0.2137</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>0.0108</td>
<td>0.4563</td>
<td>0.0208</td>
<td>0.0186</td>
<td>0.0213</td>
<td>0.3757</td>
<td>0.4035</td>
<td>0.2143</td>
<td>0.2161</td>
<td>0.4739</td>
<td></td>
</tr>
<tr>
<td>0(1)</td>
<td>0.0026</td>
<td>0.2808</td>
<td>0.0404</td>
<td>0.0656</td>
<td>0.1095</td>
<td>0.4000</td>
<td>0.2520</td>
<td>0.2237</td>
<td>0.2545</td>
<td>0.4100</td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>0.0026</td>
<td>0.2690</td>
<td>0.0161</td>
<td>0.0196</td>
<td>0.0220</td>
<td>0.4074</td>
<td>0.0658</td>
<td>0.2785</td>
<td>0.1757</td>
<td>0.1609</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>0.0042</td>
<td>0.3066</td>
<td>0.0062</td>
<td>0.0105</td>
<td>0.0189</td>
<td>0.1075</td>
<td>0.2485</td>
<td>0.0907</td>
<td>0.1019</td>
<td>0.2226</td>
<td></td>
</tr>
<tr>
<td>0(1)</td>
<td>0.0015</td>
<td>0.2205</td>
<td>0.0320</td>
<td>0.0565</td>
<td>0.0769</td>
<td>0.3681</td>
<td>0.1869</td>
<td>0.1586</td>
<td>0.2443</td>
<td>0.3422</td>
<td></td>
</tr>
<tr>
<td>200</td>
<td>0.0015</td>
<td>0.2027</td>
<td>0.0159</td>
<td>0.0194</td>
<td>0.0215</td>
<td>0.3841</td>
<td>0.0443</td>
<td>0.2582</td>
<td>0.1937</td>
<td>0.1197</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>0.0024</td>
<td>0.2257</td>
<td>0.0051</td>
<td>0.0065</td>
<td>0.0081</td>
<td>0.0635</td>
<td>0.2198</td>
<td>0.0573</td>
<td>0.0975</td>
<td>0.1731</td>
<td></td>
</tr>
</tbody>
</table>

NOTE: In moment bias section, relative moment biases with respect to actual value of the moments are recorded. This is done to gain numerical stability as higher moments of Chi-squared distribution become very large.
Table 5. Comparison between perturbation technique proposed in this paper and method-I proposed by Jones.

<table>
<thead>
<tr>
<th>Distribution</th>
<th>Sample size</th>
<th>our method MISE</th>
<th>our method Hellinger's distance</th>
<th>Jones' method MISE</th>
<th>Jones' method Hellinger's distance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standard normal</td>
<td>50</td>
<td>0.0072</td>
<td>0.2386</td>
<td>0.0351</td>
<td>0.5307</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>0.0043</td>
<td>0.1791</td>
<td>0.0212</td>
<td>0.4192</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>0.0025</td>
<td>0.1338</td>
<td>0.0126</td>
<td>0.3238</td>
</tr>
<tr>
<td>Symmetric normal mixture</td>
<td>50</td>
<td>0.0359</td>
<td>0.4502</td>
<td>0.0599</td>
<td>0.5551</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>0.0311</td>
<td>0.4158</td>
<td>0.0471</td>
<td>0.4822</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>0.0288</td>
<td>0.3969</td>
<td>0.0401</td>
<td>0.4390</td>
</tr>
<tr>
<td>Asymmetric normal mixture</td>
<td>50</td>
<td>0.0312</td>
<td>0.4213</td>
<td>0.0414</td>
<td>0.4799</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>0.0268</td>
<td>0.3775</td>
<td>0.0328</td>
<td>0.4108</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>0.0244</td>
<td>0.3478</td>
<td>0.0280</td>
<td>0.3635</td>
</tr>
<tr>
<td>Chi squared</td>
<td>50</td>
<td>0.0045</td>
<td>0.3483</td>
<td>0.0082</td>
<td>0.4359</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>0.0027</td>
<td>0.2742</td>
<td>0.0045</td>
<td>0.3288</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>0.0015</td>
<td>0.2146</td>
<td>0.0027</td>
<td>0.2504</td>
</tr>
</tbody>
</table>