Simulation Results for Wavelet-Based Density Estimators

Eric Chicken
Department of Statistics
Florida State University
Tallahassee, FL 32306

February 2, 2004

Abstract

Simulations are performed comparing various wavelet-based density estimators with each other and more traditional kernel density estimators. The wavelet estimators are based on block thresholding. Several block selection methods are examined.

Keywords: Adaptive estimation, Besov space, block thresholding, density estimation, Hölder space.
AMS 2000 Subject Classification: Primary 62G07; secondary 62G20.

1 Introduction

The simulation work given here is used to support the theoretical work found in Chicken [2] and Chicken and Cai [3]. Those manuscripts give asymptotic results for block thresholded density estimators, but [2] contains no simulation results, and [3] contains few details. The function of this technical report is to expand on the results so far provided and to describe in more detail the implementation of the estimators.

Section 2 contains information on the thresholding methods for the estimators. Section 3 contains results for a variety of simulation parameters.

2 Thresholds

In [2] and [3], the thresholds given are of the form

$$I\left(\hat{B}_{ik} > cn^{-1}\right), \tag{1}$$

where

$$\hat{B}_{ik} = \frac{1}{l} \sum_{j \in B(k)} \hat{\beta}_{i,j}^2$$

is the average of the estimated wavelet coefficients in a block $B(k)$ of length l. This is a "hard" threshold rule. Coefficients are kept unmodified if the block in which they are located has average information greater than the threshold. If not, the coefficients are all set to 0.

The constant c from (1) is useful for theoretical purposes, but it is not practical for implementation. A replacement value is needed. Since the threshold c derives from a variance-bias tradeoff, this relation will be examined further to find a suitable, implementable threshold.

The accuracy of the estimator is measured via mean integrated squared error. For each estimated coefficient, the squared bias when removing the coefficient from the estimator at a point x is $\psi_{i,j}^2(x)\beta_{i,j}^2$. When leaving this term in the estimator, the variance it contributes to the uncertainty of the model is var $(\psi_{i,j}(x)\hat{\beta}_{i,j}) = \psi_{i,j}^2(x)$ var $(\hat{\beta}_{i,j})$. We wish to retain a coefficient $\hat{\beta}_{i,j}$ if the error incurred when removing it (the squared bias) is greater than the uncertainty present if keeping it (the variance). When the n sampled points X_m are iid from density f, this is equivalent to

$$\psi_{i,j}^2(x)\beta_{i,j}^2 \quad > \quad \psi_{i,j}^2(x)\text{var }(\hat{\beta}_{ij})$$

$$
\begin{aligned}
&= \psi^2_{i,j}(x)\mathrm{var}\ \left(\frac{1}{n}\sum_{m=1}^{n}\psi_{i,j}(X_m)\right) \\
&= \psi^2_{i,j}(x)\frac{1}{n}\mathrm{var}\ \left(\psi_{i,j}(X_1)\right) \\
&= \psi^2_{i,j}(x)\frac{1}{n}\left(E[\psi^2_{i,j}(X_1)] - (E[\psi_{i,j}(X_1)])^2\right) \\
&= \psi^2_{i,j}(x)\frac{1}{n}\left(\int f(x)\psi^2_{i,j}(x)dx - \left\{\int f(x)\psi_{i,j}(x)dx\right\}^2\right) \\
&= \psi^2_{i,j}(x)\frac{1}{n}\left(\int f(x)\psi^2_{i,j}(x)dx - \beta^2_{i,j}\right),
\end{aligned}
$$

or,

$$
\beta^2_{i,j} > \frac{1}{n}\left(\int f(x)\psi^2_{i,j}(x)dx - \beta^2_{i,j}\right). \tag{2}
$$

Since the function f is unknown, the integral in (2) must be estimated. If using compactly supported wavelets, the support of ψ_{ij} is $[2^{-i}(a+j), 2^{-i}(b+j))$ if the support if ψ is, say, $[a,b)$. Since a and b are fixed constants, then for large i the support of $\psi_{i,j}$ is very small. Let f^* be some constant, say $f^* = f(x^*)$ where x^* be some point in the interior of $[2^{-i}(a+j), 2^{-i}(b+j))$. Then

$$
\int f(x)\psi^2_{i,j}(x)dx \approx f^* \int \psi^2_{i,j}(x)dx = f^*
$$

Putting this into the bias-variance comparison yields

$$
\beta^2_{i,j} > \frac{1}{n}\left(f^* - \beta^2_{i,j}\right).
$$

The value c in the theoretical threshold appears to be quite large. It depends on the maximum of the densities in a large class of functions, as well as other values which may be large. For moderate values of n, therefore, we will remove the the factor of $1/n$ from the comparison. This leaves us with

$$
\beta^2_{i,j} > f^* - \beta^2_{i,j}
$$

or,

$$
\beta^2_{i,j} > f^*/2.
$$

Since $\beta_{i,j}$ are not known, we replace them with their estimates:

$$
\hat{\beta}^2_{i,j} > f^*/2.
$$

Block thresholding argues that this comparison is more accurate if the left side is replaced with the average of several neighboring coefficients. This then becomes

$$
\hat{B}_{ik} = \frac{1}{l}\sum_{j\in B(k)}\hat{\beta}^2_{i,j} > f^*/2.
$$

The moderate value of n assumption is not unreasonable given that density estimation procedures often use small sample sizes. Indeed, for large n density estimation can be reasonably well done with simple estimators such as the histogram. In certain cases, however, this appears to make the threshold too large. See Section 3.

The point x^* may chosen to be at the center of the support of the functions $\psi_{i,j}$ corresponding to the coefficients in a particular block. The value of f at this point is determined by a pilot estimate of the density at this point. For our simulations, f was initially estimated by a simple gaussian kernel.

Instead of using just a single value of x^* in the support of the ψ, one could also use the integral, or average, of the f values in the support. For the simulations given in Section 3, the actual value used is a combination of the two: f^* is the minimum of $f(x^*)$ and the integral of f over the support interval. The estimator with this threshold will be referred to as DenBlock.

Two of the competing estimators examined come from Cai and Silverman [1]. These estimators, NeighCoeff and NeighBlock, are wavelet estimators where the comparison against a threshold is not based on a single coefficient (as is done in VisuShrink, for example) or on a single block of coefficients (as is done with DenBlock). Rather, neighboring coefficients or blocks are considered when making the threshold comparison for a particular coefficient or block. These estimator's were devised for nonparametric regression settings, but they are easily modified to the density estimation problem at hand.

For NeighBlock, the block length is $l = \log n/2$. However, the variance and squared bias used for thresholding is computed not just from the current block, but includes information from its neighbors to the immediate left and right (when possible). The total block size used for making the thresholding decision is $\log n$ when the neighboring blocks are added in. The variance of the coefficients in the extended block is replaced by the minimum of the pilot estimate of f evaluated in the center of the block and integral of f over the support of the block.

NeighCoeff is NeighBlock with block length $l = 1$. The extended block is of length 3. Again, the appropriate substitution is made in the threshold comparison as before.

A notable difference between DenBlock and the other two estimators is the thresholding rule used. DenBlock uses "hard" rule: coefficients are kept intact or set to 0. The other two use a "soft" rule: coefficients are set to 0 or shrunk toward 0. NeighBlock and NeighCoeff use the following threshold rule:

$$\left(1 - \frac{f^*}{2B^*_{i,k}}\right)_+$$

where $B^*_{i,k}$ is the sum of the squared estiamted coefficients in the extended block. This

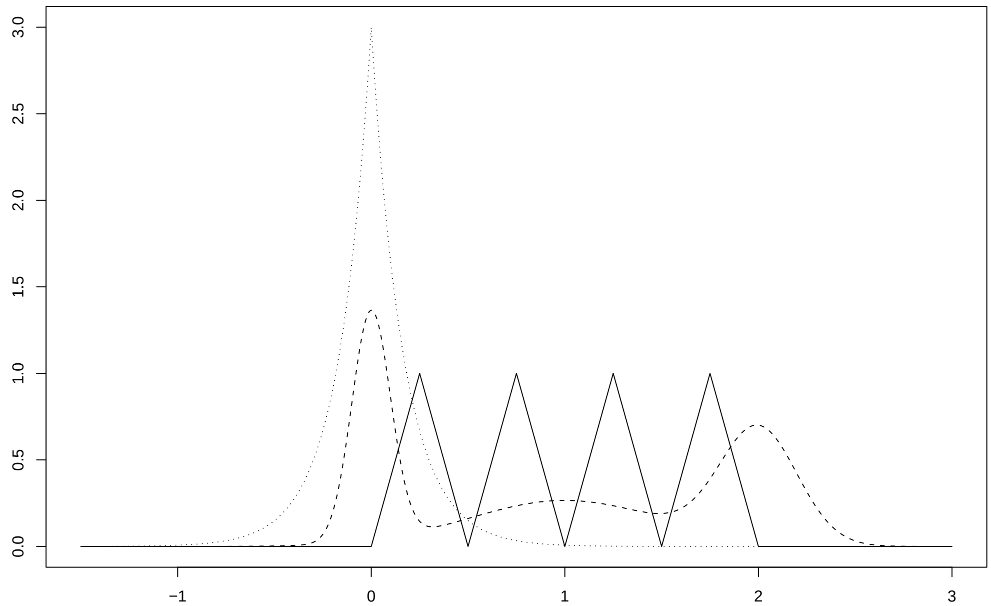

Figure 1: Test Densities. Solid line is *saw*, dashed line is *mixnorm*, and dotted line is the double exponential.

is not the exact use of the threshold found in Cai and Silverman. The form of the the threholds is the same, but the parameters are changed. NeighBlock uses a λ value in its threshold that is specifically derived from the nonparametric regression setting. Similarly for NeighCoeff. However, even in the nonparametric regression setting, the thresholds still follow from a bias-variance analysis.

For all wavelet estimators, the upper level of resolution is $\log_2(n)$. For the lower value, we use the value from Cai and Silverman (suitably modified for small n so that j_0 is always at least 1):

$$j_0 = \log_2(\log(n)).$$

This lower level is data dependent. Larger sample sizes will start the wavelet decomposition at higher reolution levels. This may also explain why for large sample sizes many of the detail coefficients are 0: the coarse approximation is capturing all important details.

Additionally, other estimators were also looked at. Biased cross-validation and unbiased cross validation kernel estimators with normal and triangle kernels were all implemented. In general, the unbiased cross-validation normal kernel (UCVN) estimator was the best of

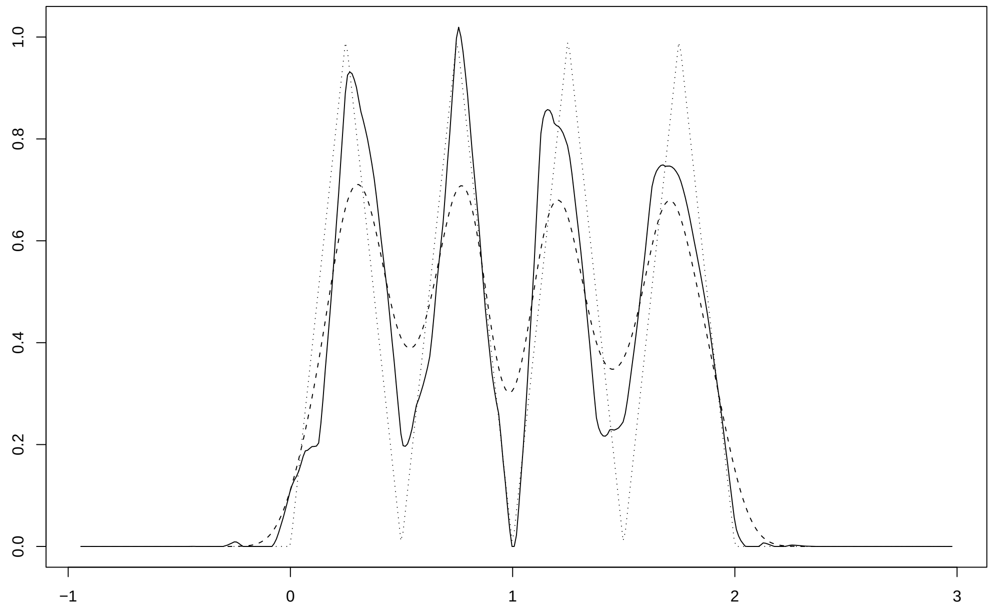

Figure 2: Typical reconstruction of *saw*. $n = 100$. Solid line is DenBlock estimate, dashed line is UCVN estimate, and dotted line is actual density.

these methods.

3 Results

These estimators were compared against one another in terms of mean squared error on the three test densities given in Figure 1. *Saw* is a comination of sums of uniform random variables, *mixnorm* is a mixture of three normal densities, and the last is a double exponential random variable. Formulas for these densities are:

$$f_{saw}(x) = \sum_{i=0}^{3} g(4x - 2i),$$

where $g(x)$ is the density for the sum of two independent uniform $(0, 1)$ random variables;

$$f_{mix}(x) = 1/3\left(10\phi(10x) + 2\phi(2(x-1)) + 5\phi(4(x-2))\right),$$

where ϕ is the density for the standard normal; and,

$$f_{dblexp} = 3h(6x) + 3h(-6x),$$

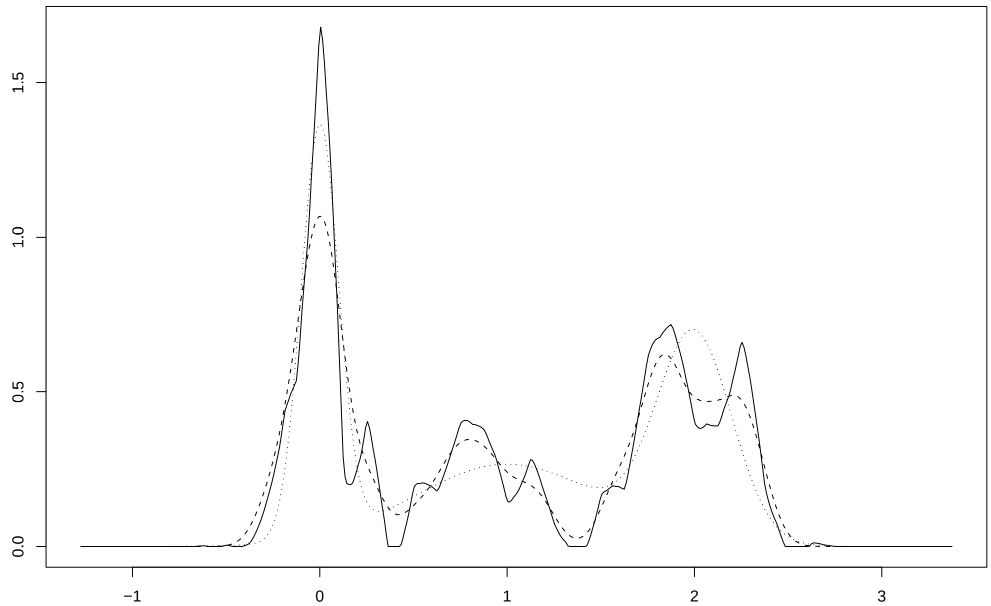

Figure 3: Typical reconstruction of *mixnorm*. $n = 100$. Solid line is DenBlock estimate, dashed line is UCVN estimate and dotted line is actual density.

where $h(x)$ is the density for an exponential random variable with mean 1.

Results of simulations for some of these estimators on various sample sizes are given in Tables 1 and 2. In each table, the MSE of the estimate is given from a repetition of size 60.

Table 1 show MSEs for samples sizes $n = 20, 50, 100, 500, 1000$ and 2000. For *saw*, the wavelet estimators have lower MSEs than UCVN with the exception of sample size 100. Once the sample size hits 100, all three of the wavelet estimators have the same MSE. Examination of the thresholded coefficients reveals that for large sample sizes, all the detail coefficients calculated are 0. Since the coarse coefficients are the same for each wavelet estimate, the wavelet estimates are identical as well. At the lower sample sizes, NeighCoeff seems preferable, then DenBlock. This agrees well with the simulation results from Cai and Silverman [1].

One possible explanation for so many zeros in the detail coefficients is that the threshold is too high. It could be lowered by adding the $1/n$ term back in, but this leads to poor results, i.e. it is too small. Simulation results (not reported here) show that with a threshold of this size, no thresholding is performed and the estimators are extremely poor. What is needed is some middle value.

For the *mixnorm* density, the UCVN is generally the best. This is perhaps not surprising given that the kernel for UCVN is from the same family as the density being estimated. Again, NeighCoeff is the best of the wavelet methods, while there is no clear distinction between NeighBlock and DenBlock.

On the final density, the double exponential, the UCVN is the worst of the estimates. The three wavelet estimators are approximately equivalent with the exception of the lowest sample size.

Similar results hold when the threshold is lowered by replacing the $1/n$ (too small) with $1/\log n$. See Table 3.

Since the wavelet estimators are approximately equivalent in terms of MSE in large sample sizes ($n \geq 50$), it is instructive to examine how the estimators work with respect to small sample sizes. The results are given in Table 2. Here, the sample sizes are $n = 10, 15, 20, 25, 30$ and 40.

For *saw*, NeighCoeff is clearly the best estimator. It is only surpassed by UCVN at the very low size $n = 10$. NeighBlock has lower MSE than DenBlock for the lower sample sizes, while DenBlock takes the lead for the larger sample sizes.

As with the sample sizes in Table 1, the UCVN is generally best at approximating *mixnorm*. NeighCoeff is the next best, while DenBlock and NeighBlock follow the same relation as they did for *saw*.

On the double exponential, NeighCoeff is clearly best over the sample sizes in Table 2, followed by NeighBlock, DenBlock, and lastly, UCVN.

Table 4 shows similar results with the threshold containing $1/\log n$ term.

The theorems in this paper show that DenBlock attains optimal convergence rates asymptotically. For the sample sizes considered here, however, NeighCoeff seems superior in terms of MSE. In particular, NeighCoeff is better than DenBlock at low sample sizes. The distinction between the wavelet estimators becomes blurred as the sample size increases. This suggests that NeighCoeff should be used for sample sizes under 50, and any of the three estimators are acceptable for larger n.

Some example reconstructions are given in Figures 2 and 3. Figure 2 shows a comparison of DenBlock and UCVN with a sample size of 100 on the *saw* density. DenBlock does well at attaining the peaks and valleys of the density. UCVN clearly shows a density with four modes, but does not capture the same highs and lows that DenBlock does. Figure 3 is a typical reconstruction of the *mixnorm* density. Here, DenBlock does a good job at estimating the peak on the left, but is too irregular over the central smooth portion. UCVN underestimates the peak, but outperforms DenBlock on the smoother portion of the density.

References

[1] T. Cai and B. Silverman. Incorporating information on neighboring coefficients into wavelet estimation. *Sankhya Ser. B*, 63:127–148, 2001.

[2] E. Chicken. Simulation results for wavelet-based density estimators. *Technical Report 01-12, Department of Statistics, Purdue University*, 2001.

[3] E. Chicken and T. Cai. Block thresholding for density estimation: local and global adaptivity. *submitted*, 2004.

Density n	DenBlock	NeighBlock	NeighCoeff	UCVN	BCVN	UCVT	BCVT
Saw							
20	10.8811	11.3950	9.15801	15.5913	16.2868	16.0360	16.2770
50	5.2177	5.5801	5.2177	7.0664	7.2917	7.2818	7.2892
100	5.2608	5.2608	5.2608	3.9586	4.4345	4.2034	4.4311
500	1.1692	1.1692	1.1692	1.4283	1.4466	1.4470	1.4466
1000	0.6061	0.6061	0.6061	0.8280	0.8409	0.8402	0.8409
2000	0.3317	0.3317	0.3317	0.4938	0.5223	0.5221	0.5223
Mixnorm							
20	23.7321	17.3795	12.7059	13.3004	14.1281	14.1562	14.1226
50	5.8260	5.8558	5.8260	6.9934	7.0133	6.8020	7.0220
100	6.0237	6.0237	6.0236	4.2097	4.2629	4.2007	4.2764
500	1.4189	1.4189	1.4189	1.2991	1.3124	1.3127	1.3125
1000	0.7400	0.7340	0.7400	0.7527	0.7656	0.7648	0.7654
2000	0.4458	0.4458	0.4458	0.4135	0.4186	0.4184	0.4184
Dbl Exp							
20	16.8195	15.0414	13.7677	20.1013	20.1092	20.1457	20.1107
50	6.9654	6.7317	6.7394	9.6706	10.1406	10.2069	10.1408
100	5.3400	5.3204	5.3203	5.3918	5.5117	6.4983	6.5128
500	1.4070	1.3979	1.3975	1.8868	1.9177	1.9253	1.9181
1000	0.9596	0.9577	0.9576	1.1357	1.1563	1.1746	1.1557
2000	0.6773	0.6763	0.6763	0.7454	0.7576	0.7530	0.7577

Table 1: MSE for Saw, Mixnorm and Double Exponential densities, threshold = $f/2$

Density n	DenBlock	NeighBlock	NeighCoeff	UCVN	BCVN	UCVT	BCVT
Saw							
10	42.8131	32.9332	25.6171	23.3807	23.7933	23.7958	23.7957
15	26.4683	21.5793	18.2141	21.6358	20.2546	20.1951	20.2512
20	11.1075	10.1879	8.0783	16.4509	17.0700	16.8357	17.0610
25	7.5533	8.7975	7.5533	13.9911	14.3080	14.1015	14.3005
30	7.1597	7.3624	7.1597	10.8101	10.9620	10.7348	10.9504
40	5.9266	6.0048	5.9266	9.1613	9.5150	9.5681	9.5039
Mixnorm							
10	51.7412	35.8192	24.5506	22.5039	22.5990	22.4803	22.6132
15	25.4060	20.9517	17.5232	17.0204	17.4138	18.1043	17.3167
20	20.2607	17.8310	14.4943	14.3062	15.1915	15.0116	15.2003
25	12.6726	12.6614	11.2159	10.2094	11.3892	11.1270	11.3922
30	8.3123	9.7426	8.3286	9.2793	9.9592	9.6868	9.9671
40	7.2358	7.8616	7.2358	8.2652	8.6300	8.3400	8.6369
Dbl Exp							
10	44.4253	29.4158	22.4899	24.6614	25.0844	25.1010	25.1011
15	27.8926	17.7815	14.9310	19.5591	20.2339	20.2091	20.2547
20	17.5577	14.0685	12.6403	17.4856	18.1337	18.1464	18.1465
25	11.4431	9.8476	9.7666	13.3411	13.5056	13.5188	13.5181
30	9.9258	9.2269	9.1675	12.9358	12.9421	12.9154	12.9538
40	8.7121	8.1441	8.1549	12.7721	12.7065	12.7067	12.7140

Table 2: MSE for Saw, Mixnorm and Double Exponential densities, threshold = $f/2$

Density n	DenBlock	NeighBlock	NeighCoeff	UCVN	BCVN	UCVT	BCVT
Saw							
20	21.4184	14.4351	8.5090	16.7697	17.1082	16.8999	17.1010
50	5.0762	5.6577	4.9466	6.7799	7.0914	6.8280	7.0709
100	5.6158	5.6158	5.6158	4.0278	4.3324	4.1121	4.3298
500	1.0561	1.0561	1.0561	1.2730	1.2924	1.2932	1.2923
1000	0.5931	0.5931	0.5931	0.8302	0.8463	0.8478	0.8461
2000	0.3257	0.3257	0.3257	0.5025	0.5104	0.5104	0.5104
Mixnorm							
20	29.2358	19.8173	12.7695	13.2680	13.6791	13.5595	13.6914
50	6.4892	6.9461	5.6884	6.7612	7.1588	6.9161	7.1679
100	6.3148	6.3148	6.3148	3.9798	4.3314	4.0717	4.3474
500	1.4123	1.4123	1.4123	1.2109	1.2460	1.2450	1.2468
1000	0.8218	0.8218	0.8218	0.7821	0.7874	0.7877	0.7873
2000	0.4376	0.4376	0.4378	0.4035	0.4090	0.4084	0.4089
Dbl Exp							
20	21.6916	15.5302	12.3255	16.5170	16.5836	16.5719	16.6024
50	7.2545	6.4270	6.5454	10.8262	11.0079	10.9884	11.0180
100	5.6471	5.5942	5.5932	5.9929	6.0391	6.0247	6.0440
500	1.3731	1.3627	1.3624	1.8715	1.8995	1.8972	1.8998
1000	0.9364	0.9341	0.9338	1.1893	1.2290	1.2308	1.2293
2000	0.6671	0.6659	0.6658	0.8912	0.9045	0.9019	0.9047

Table 3: MSE for Saw, Mixnorm and Double Exponential densities, threshold = $f/\log n$

Density n	DenBlock	NeighBlock	NeighCoeff	UCVN	BCVN	UCVT	BCVT
Saw							
10	47.2758	32.0445	22.1746	24.4620	24.8413	24.8408	24.8407
15	27.2601	23.2180	18.5890	20.9445	21.3299	21.5613	21.3270
20	20.4278	13.7675	8.7113	16.0962	16.4506	16.4453	16.4452
25	11.1824	11.7131	8.2834	13.2991	13.6412	13.6372	13.6371
30	7.8454	8.2772	7.0092	11.0690	11.4098	11.1752	11.3990
40	7.0613	7.7372	5.8905	8.0875	8.4140	8.1469	8.3945
Mixnorm							
10	53.4280	38.7518	26.3564	21.4450	21.9004	21.8242	21.9119
15	34.6479	24.3122	17.2069	16.4599	16.9084	16.7034	16.9113
20	35.1052	23.0793	14.2791	12.7021	13.1324	12.8719	13.1396
25	15.4807	16.0207	10.6528	11.2182	11.6362	11.5206	11.6494
30	13.2861	12.5759	8.8837	10.0547	10.4580	10.1422	10.4638
40	8.5148	9.1812	6.7189	7.5516	7.9783	7.7323	7.9841
Dbl Exp							
10	51.4727	31.8074	25.5616	28.3879	29.0317	29.0637	29.0626
15	31.4961	19.7367	15.6056	20.6519	20.7653	20.7676	20.7674
20	19.3912	13.5711	11.5558	14.3226	14.4451	14.7564	14.4605
25	16.5348	11.6730	10.3711	15.3616	15.6058	15.5310	15.6306
30	12.5147	9.8006	9.3896	13.7157	13.9725	13.8900	13.9945
40	8.1370	7.2189	7.0246	9.7593	10.0288	10.0282	10.0280

Table 4: MSE for Saw, Mixnorm and Double Exponential densities, threshold = $f/\log n$