Achieving Higher-Order Convergence Rates for Density Estimation with Binned Data

Michael C. MINNOTTE

A method is proposed for achieving fast ($O(n^{-8/9})$ and faster) asymptotic mean integrated squared error convergence rates in density estimation using data which is condensed to standard histogram bin counts and edges. The method involves the two elements of interpolating splines and the correct mass proportions property, in which the probability mass for the histogram bins equals exactly the fraction of the data found in those bins. Calculation of the appropriate polynomial coefficients is examined, and it is shown that the asymptotic bias of such an estimate using a spline of order p and bin width h will be $O(h^{p+1})$, resulting in optimal $AMISE$ of $O(n^{-(2p+2)/(2p+3)})$ form. Extension to the multivariate density estimation setting is described, and the effect of choice of bin edge location is briefly examined.

KEY WORDS: Histospline, Nonparametric, Smoothing

1. BINNED DATA IN NONPARAMETRIC DENSITY ESTIMATION

Among sophisticated users, the most popular method for nonparametrically estimating a probability density function f is known as kernel density estimation (Rosenblatt, 1956; Parzen, 1962). Given a sample $\{X_1, \ldots, X_n\}$ of size n, the kernel density estimate at x is computed as

$$\hat{f}_h(x) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right)$$

Michael C. Minnotte is Assistant Professor of Statistics, Department of Mathematics and Statistics, Utah State University, Logan, UT 84322-3900. The author would like to thank Daniel K. Nakano for his valuable insights into combinatoric algebra.

where K is usually a symmetric probability density function and h is a smoothing parameter known as the bandwidth. Kernel methods have been extensively studied in recent years, and perform well under many circumstances.

One limitation of the kernel density estimate is its reliance on the exact values of all data points. This can require a great deal of storage space for large data sets. It can make the calculation of the estimates quite computationally intensive. And finally, it is not always realistic. In some sense it is never realistic, since all data is collected only to some finite level of precision. A number of methods are available for overcoming these problems to various degrees. Most work with the idea of binning the data in some way.

Scott (1985a, 1992) suggests using an average shifted histogram (ASH) as an approximation to kernel estimates. We bin the data into many narrow bins, and then evaluate on grid points based on a weighted average of the bin counts. The weights can be chosen to approximate any kernel of our choice with any bandwidth. For very large data sets, this can result in a decrease in storage space requirements, and Fan and Marron (1994) demonstrate the sorts of massive speed increases which can occur through clever implementation of such ideas. It is not clear how such approximations affect the theoretical performance of kernel estimates, though usually the visual effects are quite minor, and sometimes nonexistent.

If the data is prebinned into much wider bins, however, it is questionable how useful such an approach can be. Certainly the approximation becomes progressively worse as the bins become large. In such a case it becomes worthwhile to investigate nonparametric density estimation methods specifically designed to handle binned data.

The most basic such estimator is also by far the best known nonparametric density estimation technique, the density-scale histogram. Given bin width h, and bin origin t_0, we construct a sequence of bin centers $t_k = t_0 + kh$. If the number of data points in bin

$$B_k = [t_k - \frac{h}{2}, t_k + \frac{h}{2})$$

is

$$\nu_k = \sum_{i=1}^{n} I_{B_k}(X_i),$$

then the density estimate for $x \in B_k$ will be

$$\hat{f}_{hist}(x) = \frac{\nu_k}{nh}.$$

This estimate, while simple to construct and explain, suffers theoretically relative to kernel estimates. Under the usual asymptotic mean integrated squared error criterion, the latter can achieve $O(n^{-4/5})$ convergence rates using positive (p.d.f.) kernels, and rates even closer to $O(n^{-1})$ if we relax the positivity constraint and use carefully chosen kernels with negative regions in the tails. (See, for example, Silverman, 1986, Scott, 1992, or Wand and Jones, 1995.) A histogram, on the other hand, is not so fortunate. Its piecewise constant nature prevents it from tracking even the linear term in the Taylor series of the true underlying density. Even with an optimal choice of h, the best we can hope for is $O(n^{-2/3})$ AMISE (Scott, 1979).

A partial solution to this dilemma is to linearly interpolate the histogram bin centers. The resulting frequency polygon has form

$$\hat{f}_{FP}(x) = \left(1 - \frac{x - t_k}{h}\right) \frac{\nu_k}{nh} + \left(\frac{x - t_k}{h}\right) \frac{\nu_{k+1}}{nh} \tag{1}$$

for x in $[t_k, t_{k+1}]$. The new estimate is continuous, although its derivative is not, and its linear nature allows $O(n^{-4/5})$ AMISE convergence rates, the same as positive kernels (albeit with a higher constant; Scott, 1985b).

2. CORRECT MASS PROPORTIONS AND THE BIAS-OPTIMIZED FREQUENCY POLYGON

One desirable feature of a density histogram is that the area under the density estimate between any two adjacent histogram bin edges $[t_k - h/2, t_k + h/2)$ is always

$$\frac{\nu_k}{nh} \times h = \frac{\nu_k}{n},$$

the fraction of data points found in the bin. We say that the histogram has correct mass proportions for its bins. Integration shows, however, that for a standard frequency polygon,

$$\int_{t_k-h/2}^{t_k+h/2} f^*(x)\,dx = \frac{1}{n}\left(\frac{1}{8}\nu_{k-1} + \frac{3}{4}\nu_k + \frac{1}{8}\nu_{k-1}\right).$$

The frequency polygon has incorrect mass proportions for the very histogram bins it is built from; probability mass is shifted from bins with large numbers of points to lower-density neighbors.

Minnotte (1996) describes the bias-optimized frequency polygon, which restores the correct mass proportions property to the frequency polygon. The resulting estimator remains piecewise linear between histogram bin centers, but the heights are chosen to ensure correct mass proportions. In essence, the ν_k's of equation (1) are replaced with τ_k's, calculated as

$$\tau_k = \sum_{i=-\infty}^{\infty} c_i \nu_{k+i},$$

where

$$c_i = 2^{1/2}(2^{3/2} - 3)^{|i|}.$$

This choice of weights automatically adjusts the heights at the histogram bin centers t_k so that the correct mass proportions property will apply.

We should recognize that the weights imply that somewhere, the new estimate will go negative, so that it is not a true density. Although unfortunate, this is neither unheard of, nor necessarily fatal; the higher-order kernels do this, and Gajek (1986) suggests a general modification to negative density estimates which can eliminate such regions if we so desire. In any event, the effects are usually minor and constrained to the tails of the density.

Minnotte (1996) demonstrates that the bias-optimized frequency polygon has minimal asymptotic integrated bias among piecewise linear density estimates with hinges at histogram bin centers. Despite an associated increase in the asymptotic integrated variance, the optimal bias-optimized frequency polygon represents about a 4.4% improvement in AMISE over the optimal standard frequency polygon.

3. UNIVARIATE HIGHER-ORDER HISTOSPLINES

Another view of the bias-optimized frequency polygon is that it is the linear spline function with knots at the histogram bin centers and the correct mass proportions property. If we take the above idea, but replace linear splines with ones of order p, we get the following density estimate:

$$\hat{f}_p(x) = \frac{1}{nh} \sum_{m=0}^{p} \sum_{i} b_{im} \nu_{k-i} \left(\frac{x - t_k}{h} \right)^m , \tag{2}$$

again for x in $[t_k, t_{k+1} = t_k + h]$, where the sum over i (and any other sums whose limits are not explicitly noted) will be over the integers. This estimate can trace its ancestry both to the bias-optimized freqency polygon, and to the original histospline ideas of Boneva, Kendall, and Stefanov (1971). To ensure continuity of the function and the first $p - 1$ derivatives at the knots, we require that

$$b_{im} = \sum_{j=m}^{p} \binom{j}{m} b_{(i+1)j} \qquad \forall i, \quad m = 0, \ldots, p-1. \tag{3}$$

Finally, to ensure correct mass proportions, we must have

$$\int_{t_k - \frac{h}{2}}^{t_k + \frac{h}{2}} \hat{f}_p(x)\, dx = \frac{\nu_k}{n},$$

or

$$\sum_{m=0}^{p} \frac{1}{(m+1)2^{m+1}} \left[b_{im} + (2^{m+1} - 1) b_{(i+1)m} \right] = \begin{cases} 1 & i = 0 \\ 0 & \text{otherwise} \end{cases} . \tag{4}$$

We can find the coefficients b_{im} above by first using the continuity requirements to note that

$$b_{im} = \binom{p}{m} \sum_{j=1}^{p-m} \sum_{q=1}^{j} (-1)^{j-q} \binom{j}{q} q^{p-m} \sum_{r=i+j}^{\infty} \binom{r-i-1}{j-1} b_{rp} \quad m = 0, \ldots, p-1. \tag{5}$$

This allows us to write the correct mass proportions requirement entirely in terms of infinite sums of the b_{ip} terms. By taking p differences of equation (4), we can arrive at the infinite system of equations

$$\frac{1}{(p+1)(-2)^{p+1}} \sum_{r=i-p}^{i+1} \left\{ \sum_{m=0}^{r-i+p} (-1)^{m+1} \binom{p+2}{m} [2(r-i+p-m)+1]^{p+1} \right\} b_{rp} =$$

$$\begin{cases} (-1)^i \binom{p}{i} & i = 0, 1, \ldots, p \\ 0 & \text{otherwise} \end{cases} .$$

To find the b_{ip}'s, we must solve this by inverting an infinite matrix with $p+2$ nonzero but constant diagonals, which is difficult analytically, but easily approximated by computational inversion of large matrices of the appropriate form. We can then use equation (5) to find the remaining coefficients.

This is mildly computationally intensive the first time it is done, but the coefficient calculations only need to be done once for any given value of p. Even orders do not appear to give stable solutions to the infinite matrix inversion, but odd orders do. Table 1 indicates primary values of b_{im} for $p = 1, 3,$ and 5 and nonnegative values of i. For m less than p, $b_{(-i)m} = (-1)^m b_{im}$, and for positive i, $b_{(-i)p} = -b_{(i+1)p}$.

The resulting spline-based density estimate, or "histospline," is similar to that derived by Terrell (1993) in a maximum-likelihood context, but differs in only requiring simple (histogram) binning, rather than the more sophisticated binning techniques used by Terrell.

Figure 1 shows the "kernels" associated with the linear ($p = 1$), cubic ($p = 3$), and quintic ($p = 5$) higher-order histosplines calculated from a single point. It would be possible, though inefficient, to calculate a full estimate by adding up n such curves, appropriately rescaled, with ν_k of them centered on the midpoint of each bin k.

Figure 2 shows an example of the cubic higher-order histospline for a sample of 1000 data points drawn from a standard normal density. It also shows the linear estimate (the bias-optimized frequency polygon), and the histogram from which both were constructed. Notice how, despite the poor choice of bin origin (in this case, the midpoint of the range of the data), the piecewise cubic curve does an excellent job of tracking the true density. Even the piecewise linear estimate does

	$p = 1$		$p = 3$			
i	b_{i0}	b_{i1}	b_{i0}	b_{i1}	b_{i2}	b_{i3}
0	1.414	-1.657	1.200	0.000	-3.039	1.707
1	-0.243	1.657	-0.132	0.957	2.082	-1.707
2	0.042	-0.284	0.044	-0.359	-0.766	0.949
3	-0.007	0.049	-0.016	0.130	0.277	-0.348
4	0.001	-0.008	0.006	-0.047	-0.100	0.126
5	0.000	0.001	-0.002	0.017	0.036	-0.045
6	0.000	0.000	0.001	-0.006	-0.013	0.016
7	0.000	0.000	0.000	0.002	0.005	-0.006
8	0.000	0.000	0.000	-0.001	-0.002	0.002
9	0.000	0.000	0.000	0.000	0.001	-0.001
10	0.000	0.000	0.000	0.000	0.000	0.000
	$p = 5$					
	b_{i0}	b_{i1}	b_{i2}	b_{i3}	b_{i4}	b_{i5}
0	1.180	0.000	-2.376	0.000	1.699	-0.621
1	-0.119	1.064	1.602	-0.581	-1.408	0.621
2	0.041	-0.500	-0.602	0.548	0.844	-0.450
3	-0.019	0.242	0.279	-0.290	-0.425	0.254
4	0.009	-0.118	-0.135	0.143	0.209	-0.127
5	-0.004	0.058	0.066	-0.070	-0.102	0.062
6	0.002	-0.028	-0.032	0.034	0.050	-0.030
7	-0.001	0.014	0.016	-0.017	-0.024	0.015
8	0.001	-0.007	-0.008	0.008	0.012	-0.007
9	0.000	0.003	0.004	-0.004	-0.006	0.004
10	0.000	-0.002	-0.002	0.002	0.003	-0.002
11	0.000	0.001	0.001	-0.001	-0.001	0.001
12	0.000	0.000	0.000	0.000	0.001	0.000
13	0.000	0.000	0.000	0.000	0.000	0.000

Table 1: Values of b_{im} for p =1, 3, and 5 to 3 decimal places.

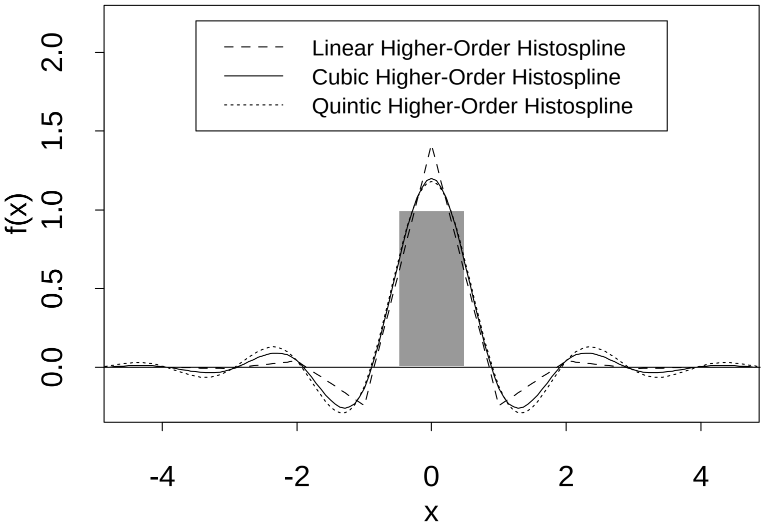

Figure 1: Linear, cubic, and quintic histosplines based on a single point in the bin centered on 0. Bin width $h = 1$ is used in all cases.

quite well, except at the mode where it is unable to follow the strong curve of the true density. The quintic estimate is nearly indistinguishable from the cubic one.

4. MEAN INTEGRATED SQUARED ERROR

The advantages of choosing our binned density estimate in this way become clear as we investigate the asymptotic squared bias properties. The only random element in our estimate as described in equation (2) is the set of bin counts, $\boldsymbol{\nu}$, a multinomial random variable, with probabilities equal to the probability of a single random variable from f falling in each histogram bin.

We can approximate the bin probabilities using a Taylor series on the underlying density f. Defining $p_{k+i} = P\left(X \in B_{k+i}\right)$, we have

$$\begin{aligned} p_{k+i} &= \int_{t_k+(i-1/2)h}^{t_k+(i+1/2)h} \left[\sum_{l=0}^{\infty} \frac{(x-t_k)^l}{l!} f^{(l)}(t_k) \right] dx \\ &= \sum_{l=0}^{\infty} h^{l+1} \left(\sum_{j=0}^{\lfloor l/2 \rfloor} \frac{i^{l-2j}}{(2j+1)!(l-2j)!2^{2j}} \right) f^{(l)}(t_k) \end{aligned}$$

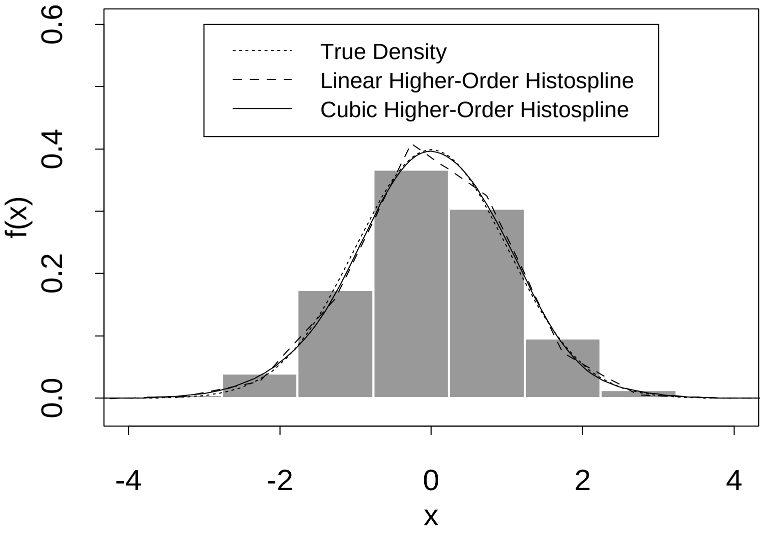

Figure 2: Histogram and linear and cubic higher-order histospline estimates of a $N(0,1)$ density (also shown). Bin width $h=1$ is used for all estimates.

where $\lfloor z \rfloor$ is the greatest integer no greater than z. With these probabilities, we can calculate (under the assumption that $x \in [t_k, t_{k+1}]$)

$$E[\hat{f}_p(x)] = \sum_{l=0}^{p} \sum_{m=0}^{p} \frac{h^{l-m}(x-t_k)^m}{(l+1)!} f^{(l)} \sum_{j=0}^{\lfloor l/2 \rfloor} \left[\binom{l+1}{2j+1} \frac{1}{2^{2j}} \sum_i i^{l-2j} b_{im} \right] + O(h^{p+1}).$$

Since the Taylor series of f around t_k has form

$$f(x) = \sum_{l=0}^{p} \frac{(x-t_k)^l}{l!} f^{(l)} + O\left((x-t_k)^{p+1}\right),$$

the asymptotic bias will be $O(h^{p+1})$ if

$$\frac{h^{l-m}}{l+1} \sum_{j=0}^{\lfloor l/2 \rfloor} \binom{l+1}{2j+1} \frac{1}{2^{2j}} \sum_i i^{l-2j} b_{im} = \begin{cases} 1 & l = m \\ 0 & \text{otherwise, } 0 \le l \le p \end{cases}. \tag{6}$$

Equation (3) allows us to see that for $m = 0, \ldots, p-1$,

$$\begin{aligned} \sum_i b_{im} &= \sum_{j=m}^{p} \binom{j}{m} \sum_i b_{(i+1)j} \\ &= \sum_{j=m}^{p} \binom{j}{m} \sum_i b_{ij}. \end{aligned}$$

If we set $m = p-1$, this leaves us with $\sum_i b_{ip} = 0$. Working down from $p-1$, similar calculations show that $\sum_i b_{im} = 0$ for $m = 1, \ldots, p$.

Summing both sides of equation (4) over all i, and discarding all terms now known to be 0, we find

$$\frac{1}{2}\left(\sum_i b_{i0} + \sum_i b_{(i+1)0}\right) = 1,$$

and therefore $\sum_i b_{i0} = 1$.

Likewise, we can say that

$$\begin{aligned}
\sum_i i b_{im} &= \sum_{j=m}^{p} \binom{j}{m} \sum_i i b_{(i+1)j} \\
&= \sum_{j=m}^{p} \binom{j}{m} \sum_i (i-1) b_{ij} \\
&= \sum_{j=m}^{p} \binom{j}{m} \left(\sum_i i b_{ij} - \sum_i b_{ij}\right).
\end{aligned}$$

Again, starting with $m = p-1$ and working down, we get

$$\sum_i i b_{im} = \begin{cases} 0 & m = 2, \ldots, p \\ 1 & m = 1 \end{cases},$$

and equation (4) will allow us to verify that $\sum_i i b_{i0} = 0$.

Continuing in this fashion, working inductively up on the powers of i and down on m, we can show that for $0 \leq l \leq p$,

$$\sum_i i^l b_{im} = \begin{cases} (l! T_j)/(m! 2^{2j}) & l - m = 2j, \qquad j \in \{0, 1, 2, \ldots\} \\ 0 & \text{otherwise} \end{cases},$$

where

$$T_j = \begin{cases} 1 & j = 0 \\ -\sum_{r=0}^{j-1} (T_r)/(2j - 2r + 1)! & \text{otherwise} \end{cases}.$$

A little algebra will verify that these sums are the solutions to condition (6). Any density estimate of the form described by equations (2), (3), and (4) will have $O(h^{p+1})$ asymptotic bias.

The leading term in the remaining integrated squared bias formulation will have form

$$AISB(\hat{f}_p) = \frac{h^{2p+2} \int f^{(p+1)}(x)^2\, dx}{[(p+2)!]^2} \sum_{m_1=0}^{p+1} \sum_{m_2=0}^{p+1} \frac{c_{m_1} c_{m_2}}{(m_1+m_2+1)},$$

where

$$c_m = \begin{cases} \sum_{j=0}^{\frac{p+1}{2}} \binom{p+2}{2j+1} 2^{-2j} \sum_i i^{p+1-2j} b_{im} & m = 0, \ldots, p \\ -(p+2) & m = p+1 \end{cases}.$$

The constant is (1/720) when $p = 1$, (1/1209600) when $p = 3$, and (691/259459200) for $p = 5$.

Turning to the variance, since equation (2) can be written

$$\hat{f}_p(x) = \frac{1}{nh} \sum_i \left[\sum_{m=0}^{p} b_{im} \left(\frac{x - t_k}{h} \right)^m \right] \nu_{(k-i)},$$

we can find that

$$\mathrm{Var}(\hat{f}_p(x)) = \frac{1}{n^2h^2} \sum_i \sum_j \left[\sum_{m_1=0}^{p} b_{im_1} (x - t_k)^{m_1} \right] \left[\sum_{m_2=0}^{p} b_{jm_2} (x - t_k)^{m_2} \right] \sigma_{(k-i)(k-j)}$$

with

$$\sigma_{ij} = \begin{cases} np_i(1-p_i) \approx nhf(t_i)(1 - hf(t_i)) & i = j \\ -np_ip_j \approx -nh^2 f(t_i) f(t_j) & i \neq j. \end{cases}$$

If h goes to 0 as the sample size becomes infinite, then asymptotically, the h^2 terms (the covariances and the second term of the variances) will be dominated by the $nhf(t_i)$ term of the variances. We can say that the asymptotic variance of $\hat{f}_p(x)$ is

$$AV(\hat{f}_p(x)) = \frac{1}{nh^2} \sum_i hf(t_{k-i}) \left[\sum_{m=0}^{p} b_{im} \left(\frac{x - t_k}{h} \right)^m \right]^2.$$

We find the asymptotic integrated variance by integrating over x from t_k to t_{k+1}, and then summing over the bins k, to get

$$\begin{aligned} AIV(\hat{f}_p) &= \sum_k \frac{1}{nh^2} \sum_i hf(t_{k-i}) h \sum_{m_1=0}^{p} \sum_{m_2=0}^{p} \frac{b_{im_1} b_{im_2}}{(m_1+m_2+1)} \\ &= \frac{1}{nh} \sum_i \sum_{m_1=0}^{p} \sum_{m_2=0}^{p} \frac{b_{im_1} b_{im_2}}{(m_1+m_2+1)} \sum_k hf(t_{k-i}) \\ &= \frac{1}{nh} \sum_i \sum_{m_1=0}^{p} \sum_{m_2=0}^{p} \frac{b_{im_1} b_{im_2}}{(m_1+m_2+1)} \end{aligned}$$

since the sum $\sum_k hf(t_{(k-i)})$ is simply the standard Riemannian approximation to the integral of f, and is therefore approximately 1. The constant on $(nh)^{-1}$ is 1.179, 1.209, and 1.246 for $p = 1$, 3, and 5, respectively.

The asymptotic variance is $O((nh)^{-1})$ (as is true for histograms, frequency polygons, and kernel estimates of all orders), so the optimal bin width will be $O(n^{-1/(2p+3)})$ and the optimal $AMISE$ will be $O(n^{-(2p+2)/(2p+3)})$. The constants on the $AMISE$ for $p = 1$, 3, and 5 are 0.505, 0.354, and 0.599 respectively. These are the same convergence rates found in the higher-order kernel estimates of Bartlett (1963) (using a kernel of order $p + 1$), but without the necessity of storing and computing with all of the original data points!

Although these results get better the higher the order of the polynomial, the practical improvements may be quite asymptotic once you go beyond $p = 3$. The estimates are visually nearly identical, and may share with higher-order kernel estimates the fact that by the time the sample is large enough to show large relative gains, the absolute error may be so small as to make the improvement unnecessary. For most practical applications, the cubic appears to be the most appropriate choice of polynomial.

5. ALGORITHMIC COMPLEXITY

Once the b_{im} coefficients are calculated (which need be done only once per value of p), there are two steps remaining in producing the final estimate. The binning requires $O(n)$ steps. Of course, these steps are very simple ones, and may very well be done as the data is collected.

Once the g bins are prepared, computation of the estimate theoretically requires $O(g^2)$ steps. Practically, if we eliminate the effects of bins further than some distance away (as they rapidly damp toward 0), we can find our estimate in $O(g)$ steps, all of which are simple polynomial calculations.

If we assume a fixed interval of interest and an optimal $O(n^{-1/(2p+3)})$ bin width, g will be $O(n^{1/(2p+3)})$ and the computation will be $O(n^{2/(2p+3)})$ theoretically and

$O(n^{1/(2p+3)})$ practically. This technique should remain computationally feasible for even very large data sets, where the time to bin will be the biggest constraint.

6. MULTIVARIATE HIGHER-ORDER HISTOSPLINES

The advantageous properties of the higher-order histospline can be achieved in multiple dimensions as well. Following the linear blend frequency polygon ideas of Terrell (1983) and Hjort (1986), though no longer merely linear in any sense, in d dimensions we can achieve an $AMISE$ rate of $O(n^{-(2p+2)/(2p+2+d)})$.

For example, in two dimensions, partition the $x - y$ plane into two dimensional bins

$$B_{k_x k_y} = [t_{k_x} - \frac{h_x}{2}, t_{k_x} + \frac{h_x}{2}) \times [t_{k_y} - \frac{h_y}{2}, t_{k_y} + \frac{h_y}{2}),$$

and let

$$\nu_{k_x k_y} = \sum_{i=1}^{n} I_{B_{k_x k_y}}(X_i, Y_i).$$

Define the two-dimensional order-p histospline as

$$\hat{f}_p(x, y) = \frac{1}{nh_x h_y} \sum_{m_x=0}^{p} \sum_{m_y=0}^{p} \sum_i \sum_j b_{im_x} b_{jm_y} \nu_{(k_x-i)(k_y-j)} \left(\frac{x - t_{k_x}}{h_x}\right)^{m_x} \left(\frac{y - t_{k_y}}{h_y}\right)^{m_y},$$

where the b terms are exactly those of the univariate histospline of order p. It is straightforward to show that the expectation factors in exactly the same way as the two-dimensional Taylor series on $f(x, y)$. Thus the resulting bias will have no terms of lower order than $(p + 1)$, just as in the univariate case. The asymptotic variance will now be of order $O((nh_x h_y)^{-1})$, and the $AMISE$ rate will be the one promised above.

7. CHOICE OF BIN ORIGIN

A final aspect of the higher-order histosplines worth mention is an apparent reduced dependence on the choice of bin origin, t_0. Simonoff (1995) and Simonoff and Udina (1995) have looked at the problem for histograms and standard frequency

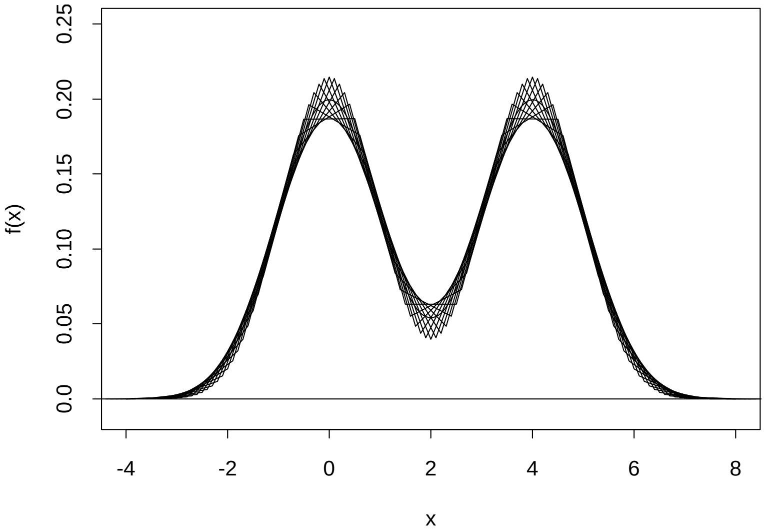

Figure 3: Expected values of 10 bias-optimized frequency polygon estimates of $0.5N(0,1) + 0.5N(4,1)$. Binwidth $h = 1$ is used for all estimates. Bin origin varies from $t_0 = 0$ to $t_0 = 0.9$ in steps of 0.1.

polygons. The latter paper provides a "stability index" assessing the change in appearance for an estimate from a given data set and bin width as bin origin changes.

Examination of the expected values of the bias-optimized frequency polygon (the higher-order histospline of order 1) and of the order-3 (cubic) higher-order histospline (Figures 3 and 4) for 10 different choices of bin origin shows that pointwise, the order-3 estimate is far less dependent on t_0. Although no formal investigation of this phenomenon has been conducted yet, it indicates both another reason for interest in the higher-order estimates, and a potential subject for further study.

REFERENCES

Bartlett, M.S., (1963), "Statistical Estimation of Density Functions," *Sankhyā Series A*, **25**, 245–254.

Boneva, L., Kendall, D., and Stefanov, I., (1971), "Spline Transformations: Three New Diagnostic Aids for the Statistical Data Analyst," *Journal of the Royal*

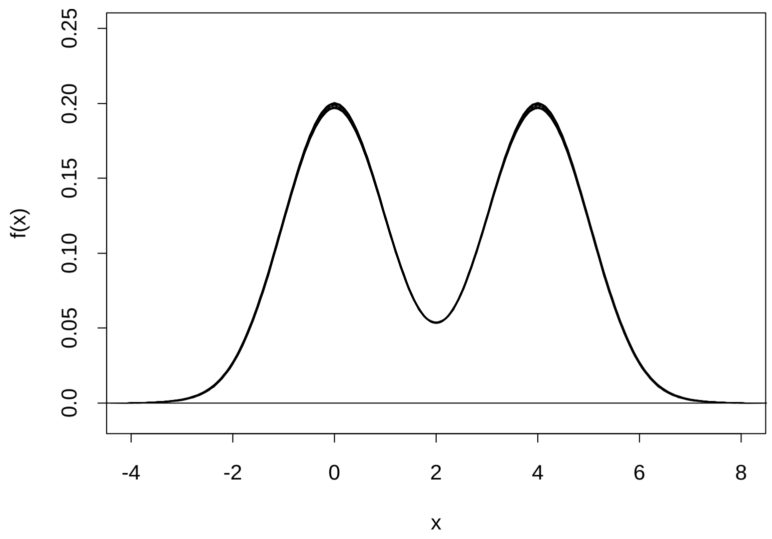

Figure 4: Expected values of 10 cubic higher-order histospline estimates of $0.5N(0,1) + 0.5N(4,1)$. Binwidth $h = 1$ is used for all estimates. Bin origin varies from $t_0 = 0$ to $t_0 = 0.9$ in steps of 0.1.

Statistical Society, **33**, 1–70.

Fan, J. and Marron, J.S., (1994), "Fast Implementation of Nonparametric Curve Estimators," *Journal of Computational and Graphical Statistics*, **3**, 35–56.

Gajek, L., (1986), "On Improving Density Estimators Which Are Not Bona Fide Functions," *The Annals of Statistics*, **14**, 1612–1618.

Hjort, N.L., (1986), "On Frequency Polygons and Average Shifted Histograms in Several Dimensions", Technical Report 22, Stanford University.

Minnotte, M.C., (1996), "The Bias-Optimized Frequency Polygon," *Computational Statistics*, **11**, 35–48.

Parzen, E., (1962), "On Estimation of Probability Density Function and Mode," *Annals of Mathematical Statistics*, **33**, 1065–1076.

Rosenblatt, M., (1956), "Remarks on Some Nonparametric Estimates of a Density Function," *Annals of Mathematical Statistics*, **27**, 832–837.

Scott, D.W., (1979), "On Optimal and Data-Based Histograms," *Biometrika*, **66**, 605–610.

Scott, D.W., (1985a), "Average Shifted Histograms: Effective Nonparametric Density Estimators in Several Dimensions," *The Annals of Statistics*, **13**, 1024–1040.

Scott, D.W., (1985b), "Frequency Polygons," *Journal of the American Statistical Association*, **80**, 348–354.

Scott, D.W., (1992), *Multivariate Density Estimation: Theory, Practice, and Visualization*, New York: John Wiley.

Silverman, B.W., (1986), *Density Estimation for Statistics and Data Analysis*, New York: Chapman and Hall.

Simonoff, J.S., (1995), "The Anchor Position of Histograms and Frequency Polygons: Quantitative and Qualitative Smoothing," *Communications in Statistics — Simulation and Computing*, **24**, in press.

Simonoff, J.S. and Udina, F., (1996), "Measuring the Stability of Histogram Appearance When the Anchor Position is Changed," *Computational Statistics and Data Analysis*, **21**, under revision.

Terrell, G.R., (1983), "The Multilinear Frequency Spline", Technical Report 85-7, Department of Math Sciences, Rice University.

Terrell, G.R., (1993), "Spline Density Estimators," *Proceedings of the Section on Statistical Computing, American Statistical Association*, , 255–260.

Wand, M.P. and Jones, M.C., (1995), *Kernel Smoothing*, New York: Chapman and Hall.