行政院國家科學委員會專題研究計畫 成果報告

一般化頻率多邊形圖的加權函數
研究成果報告(精簡版)

計 畫 類 別 ：個別型
計 畫 編 號 ：NSC 97-2118-M-032-011-
執 行 期 間 ：97 年 08 月 01 日至 98 年 07 月 31 日
執 行 單 位 ：淡江大學數學系

計 畫 主 持 人 ：伍志祥

處 理 方 式 ：本計畫可公開查詢

中 華 民 國 99 年 01 月 06 日

A Study of Frequency Polygons Based on Weighted Averages of Binned Data

ABSTRACT

We revisit the generalized midpoint or edge frequency polygons of Scott (1985b), and Dong and Zheng (2001). Their estimators are linear interpolants of the appropriate values above the bin centers or edges, those values being the weighted averages of the heights of r, $r \in N$, neighboring histogram bins. For small and moderate values of r, we obtain the optimal choices of weights that minimize the asymptotic mean integrated square error (AMISE) of their generalized frequency polygons. The minimum AMISE decreases as the values of r increases. A simple kernel evaluation method based on the truncated Epanecknikov kernel is proposed to generate the weights for binned values. The proposed method can provide near-optimal weights. In addition, we prove that the discrete uniform weight function minimizes the variance of the generalized frequency polygon under some mild conditions. Simulation study demonstrates that, as the value of r increases, the improvements in AMISE performance carry over to small and moderate sample situations.

Key words: Midpoint frequency polygons; Edge frequency polygons; Kernel-based weight; Mixture weight; Minimum variance weights.

AMS Subject Classifications: 62G07; 62G20;

1. Introduction

Given a univariate random sample of size n from an unknown continuous distribution, consider the problem of estimating the density function f by variants of histogram. The frequency polygon, e.g. Scott (1985a), is a widely used density estimator based on histogram, with some form of linear interpolation. It enjoys the computational simplicity of histogram and the statistical efficiency of nonnegative kernel density estimators. See also Scott (1992) for excellent reviews of such estimator.

Scott (1985b) averages over a collection of histograms using different anchor positions to construct a histogram-like estimator known as average shifted histogram and makes it continuous by linear interpolation. The resultant estimator known as the frequency polygon of average shifted histogram (FP-ASH) connects with straight lines the weighted average of $r = 2q - 1 (q \in N)$, say, binned values above the bin centers. The FP-ASH can suppress the noise effect of histogram's anchor shifting, while retaining many of the computational advantage of a density estimate based on bin count. We term FP-ASH as a generalized version of midpoint frequency polygons in this article. On the other hand, Dong and Zheng (2001) extends the idea of Jones et al. (1998) and proposes a generalized version of edge frequency polygon (GEFP). GEFP connects with straight lines the appropriate values above the bin edges. Those values are the weighted averages of the heights in the neighboring $r = 2q$ bins. They obtained optimal weights to minimize the asymptotic mean integrated square error (AMISE) for the cases of $r = 4$ and 6, and suggested the use of subinterval integration of kernel function to generate weights in practice.

In this article, we shall revisit the above generalized versions of midpoint and edge frequency polygons. Their definitions are given in Section 2. Call the above two

estimators simply frequency polygon if there is no danger of confusion in what follows. Given a bin width and the value of r, it is the choice of weights that determines the performance of the resultant frequency polygon. Except for the case of $r = 4$ and 6, the AMISE-optimal weights for $r \geq 3$ are unavailable and the association between the value of r and the AMISE of frequency polygon using such weights remains unknown. On the other hand, it is natural and straightforward to use kernel function to generate the weights for binned values in practice such that the resultant weights have better interpretation (see Dong and Zheng (2001)). It is desirable to have a kernel method to closely approximate the optimal choice of weights. Motivated by the above, in this article we shall (i) provide the optimal weights for some small and moderate values of r and then compare the associated AMISE performance of the frequency polygons, and (ii) propose a kernel method to generate near optimal choices of weights with computational simplicity. In addition, it is a common practice to assigns uniform weights to binned values for simplicity. A question arises here: does the use of uniform weights reduces or minimizes the variance of the resultant frequency polygon? For even values of r, Dong and Zheng (2001) has shown that the AMISE of the frequency polygon based on uniform weights is minimized when $r = 6$. In this article, the frequency polygon based on uniform weights is further investigated through its AMISE and asymptotic integrated variance (AIV).

In Section 2, we give the construction of frequency polygon and apply the formula (3.2) of Jones (1989) to obtain the AMISE of the frequency polygon for the all values of r. In Section 3, we first obtain the AMISE-optimal weights for $r = 3, 4, 5, 6, 7$ and show that the AMISE of the frequency polygon using the optimal weights monotonically decreases as the value of r increases from 2 to 7. Secondly, we tailored

the Epanecknikov kernel function through truncation to generate the weight in practice. Our proposed method can provide near-optimal weights for $r = 3, 4, 5, 6, 7$. In addition, let g be any given weight function, then the frequency polygon based on the mixture of g and uniform weight function has smaller AIV than that based on g. Moreover, for any fixed value of r, the use of minimum weight leads to minimum AIV under some mild conditions and the AMISE is minimized when $r = 5$.

A simulation study is carried out in Section 4 to compare the performances of the frequency polygons by their simulated mean integrated square error in finite sample situations, for $r = 2, 3, 4, 5, 6$ and 7. The results show that, as the value of r increases, the improved accuracy of the frequency polygon in terms of the AMISE carries over to the situations of small and moderate sample sizes.

2. Generalized midpoint and edge frequency polygons

Let $X_1, X_2, \cdots, X_n$ be a random sample from a continuous distribution with density function f. For each frequency polygon in this article, divide the sample space into equal length intervals, or bins, of length or bin width b. Let $s_k = k \cdot b$, $k \in Z$, denote the bin edges. Then $t_k = s_k + b/2$ is the bin center of the bin $B_k = [s_k, s_{k+1})$ and $n_k = \sum_{i=1}^{n} I_{B_k}(X_i)$ denotes the bin count of B_k, where I is the indicator function, namely, $I_B(x) = 1$, if $x \in B$ and 0, otherwise. Note that $t_{k+1} - t_k = b$ and $\sum_k n_k = n$ by assumption.

We now give the definition of frequency polygon as follows. Here and throughout this article, let g be a nonnegative weight function defined on $D = \{1, 2, \cdots, r\}$, with $g(i) = w_i$ and $\sum_{i=1}^{r} w_i = 1$. Define the weighted average of r neighboring bin heights for odd and even values of r as, respectively, $u_k = (1/nb)\sum_{j=1}^{r} w_j n_{k+j-(r+1)/2}$ and

$v_k = (1/nb)\sum_{j=1}^{r} w_j n_{k+j-(r+2)/2}$. The generalized midpoint and edge frequency polygons are constructed by connecting the averaged ordinates $\{t_k, u_k\}$ and $\{s_k, v_k\}$, respectively, with straight lines. That is, they are respectively defined as:

$$\begin{aligned} \hat{f}_{\mathrm{MFP},r}(x) &= \frac{t_{k+1}-x}{b}u_k + \frac{x-t_k}{b}u_{k+1},\ x \in [t_k, t_{k+1}), \\ \hat{f}_{\mathrm{EFP},r}(x) &= \frac{s_{k+1}-x}{b}v_k + \frac{x-s_k}{b}v_{k+1},\ x \in [s_k, s_{k+1}), \end{aligned} \tag{2.1}$$

for odd and even values of r.

When r is even, $\hat{f}_{\mathrm{EFP},r}$ is the GEFP considered by Dong and Zheng (2001). For the case of $r=2$, $\hat{f}_{\mathrm{EFP},2}$ uses $w_1 = w_2 = 1/2$ and is the edge frequency polygon (EFP) considered by Jones et al. (1998). The authors pointed out that the EFP $\hat{f}_{\mathrm{EFP},2}$ has a potentially better resolution of peaks while the traditional midpoint frequency polygon $\hat{f}_{\mathrm{MFP},1}$ is much more hit-and-miss. Simonoff and Udina (1997) demonstrate that the appearance of $\hat{f}_{\mathrm{EFP},2}$ is less sensitive to the choice of anchor position. On the other hand, when r is odd, $\hat{f}_{\mathrm{MFP,r}}$ is the generalized version of midpoint frequency polygon. Scott (1985b) average over q histograms of different bin origins and common bin width, h say, to obtain a new histogram and made it continuous by linear interpolation. The resultant estimator, FP-ASH, is equivalent to the midpoint frequency polygon $\hat{f}_{\mathrm{MFP,r}}$, $r = 2q-1$, using narrower bin width $b = h/q$ and the weights generated by the Triangular kernel function $w_j = 1 - |j-q|/q,\ j = 1, 2, \cdots, 2q-1,$ in our notations. We now drop the subscripts MFP and EFP from $\hat{f}_{\mathrm{MFP},r}(x)$ and $\hat{f}_{\mathrm{EFP},r}(x)$, respectively, and use $\hat{f}_r$ henceforth to denote the above two estimators if no confusion occurs in what follows.

Following the approach of Jones et al. (1998) and Dong and Zheng (2001), we shall use the AMISE of $\hat{f}_r,\ r \geq 2$ to understand the effect of the above generalization

of frequency polygons. The result of Jones (1989) is applicable for this purpose when u_k and v_k are rewritten in form of kernel estimators. We reformulate them as follows. Let

$$K_r(x) = \begin{cases} \sum_{j=1}^{r} \frac{r-1}{2} w_j I_{[(2j-r-2)/(r-1),(2j-r)/(r-1))}(x), & r \text{ is odd,} \\ \sum_{j=1}^{r} \frac{r}{2} w_j I_{[(2j-r-2)/r,(2j-r)/r)}(x), & r \text{ is even.} \end{cases} \tag{2.2}$$

It can be shown that K_r is a symmetric probability density function symmetric about 0, and u_k and v_k can be rewritten in form of kernel estimators as $u_k = \frac{1}{nb(r-1)/2} \sum_{i=1}^{n} K_r(\frac{X_i - t_k}{b(r-1)/2})$ and $v_k = \frac{1}{nbr/2} \sum_{i=1}^{n} K_r(\frac{X_i - s_k}{br/2})$, using bin width $b(r-1)/2$ and $br/2$, respectively. Proposition 2.1 gives the integrated mean square error of $\hat{f}_r(x)$, as a direct consequence of Jones (1989).

Proposition 2.1 Suppose that f'' is continuous with $R(f'') = \int f''(x)^2 dx < \infty$ and let g be a nonnegative weight function defined on D. If $n \to \infty$, $b \to 0$ and $nb \to \infty$, then the mean integrated square error of $\hat{f}_r$, $r \geq 2$, is given by:

$$\text{MISE}(\hat{f}_r) = \frac{R(f'')b^4}{4}[\mu^2 + \frac{\mu}{3} + \frac{1}{30}] + \frac{1}{3nb}[2R(K_r) + R^*(K_r)] + o(\frac{1}{nb} + b^4), \tag{2.3}$$

where $\mu = c_r^2 \int x^2 K_r(x) dx = \sum_{i=1}^{r} g(i)[(i-(r+1)/2)^2 + 1/12]$, $R(K_r) = c_r^{-1} \int K_r(x)^2 dx = \sum_{i=1}^{r} g(i)^2$, $R^*(K_r) = c_r^{-1} \int K_r(x) K_r(x - c_r^{-1}) dx = \sum_{i=1}^{r-1} g(i)g(i+1)$, $c_r = (r-1)/2$ and $r/2$, respectively, for odd and even values of r.

The leading terms, namely AMISE, are the sum of asymptotic integrated bias and AIV. The AMISE of $\hat{f}_r$ depends on the weights $w_i, i = 1, 2, \cdots, r$. Given a value of r, we need to determine the weights w_i such that the AMISE is minimized. We shall derive the optimal weights w_i for $r = 3, 4, 5, 6$ and 7 that serve the above purpose. By

(2.3), the AMISE-optimal bin width for a given value of r is

$$b_r^* = \left(\frac{\frac{2}{3}R(K_r) + \frac{1}{3}R^*(K_r)}{nR(f'')[\mu^2 + \frac{1}{3}\mu + \frac{1}{30}]} \right)^{1/5}, \quad (2.4)$$

which can be derived for the following minimum AMISE of $\hat{f}_r$, $r \geq 2$.

$$\text{AMISE}^*(\hat{f}_r) = \frac{5}{4}\left(\frac{R(f'')}{n^4} \right)^{1/5} C(r), \quad (2.5)$$

where $C(r) = [\mu^2 + \mu/3 + 1/30]^{1/5}[2R(K_r)/3 + R^*(K_r)/3]^{4/5}$, $r \geq 2$. For simplicity of notation, from now on, we use AMISE^* to denote $\text{AMISE}^*(\hat{f}_r)$.

3. The weights

By (2.5), one can choose the weights w_i to minimize $C(r)$ and thus AMISE^*. In this section we shall give the optimal weight for some values of r obtained through exhaustive grid search and propose a simple kernel evaluation method to generate weights for all values of r in practice. We shall use slightly different notations to denote $\hat{f}_r$ based on different weighing schemes. To alleviate confusion, Table 3.1 lists other notations concerning the frequency polygon $\hat{f}_r$ defined as (2.1), together with the weights used in $\hat{f}_r$ and the sections (in parenthesis) in which they are defined.

[Insert Table 3.1 about here]

3.1 Optimal weights

By (2.5), we use numerical calculation to obtain optimal weights, denoted by w_i^*, such that $C(r)$ and thus AIMSE^* are minimized. Our AMISE-optimal weights obtained through exhaustive grid search are reported as follows. For r =3, $w_1^* = w_3^* = 0.310242$, $w_2^* = 0.379517$, and when r =5, w_1^* = w_5^* = 0.150733,

$w_2^* = w_4^* = 0.223524$ and $w_3^* = 0.251485$. Finally, for $r = 7$, $w_1^* = w_7^* = 0.088860$, $w_2^* = w_6^* = 0.141709$, $w_3^* = w_5^* = 0.176049$, $w_4^* = 0.186764$. The above w_i^*'s are reported in the column 2 of Table 3.2. Let $\hat{f}_r^*$ be the frequency polygon $\hat{f}_r$ using the optimal weight function g^* defined as $g^*(i) = w_i^*$, $i = 1, 2, \cdots, r$. The $C(r)$ values associated with the above optimal weights are plotted as plus signs in Figure 3.1. Note that the optimal weights for $r = 2$ and 4 reported in Table 3.2 confirm the result of Dong and Zheng (2001).

Jones et al. (1998) has shown that $b_2^* = 0.6875 \cdot b_1^*$ and $\text{AMISE}^*(\hat{f}_2^*) = 0.8946 \cdot \text{AMISE}^*(\hat{f}_1)$, i.e. that $\hat{f}_2$ is more than 10% better than $\hat{f}_1$ in terms of AMISE. Through a straightforward calculation based on (2.4), the optimal bin widths of $\hat{f}_r^*$ has the relation $b_r^* = k_r \cdot b_2^*$, where for $r = 3, 4, 5, 6$ and 7, $k_r = 0.80004$, 0.65940, 0.55603, 0.47829 and 0.41840, respectively. It follows from (2.4) and (2.5) that for any value of b_2^*, taking $b_r^* = k_r \cdot b_2^*$ yields $\text{AMISE}^*(\hat{f}_r^*) = a_r \cdot \text{AMISE}^*(\hat{f}_2)$, where $a_r =$ 0.964256, 0.943120, 0.941186, 0.936503 and 0.933500, respectively, for $r = 3$, 4, 5, 6 and 7. So, the AMISE^* decreases as the value of r increases but the improvement levels out when r increases further. Note here the values of a_4 and a_6 confirm the result of Dong and Zheng (2001).

[Insert Table 3.2 about here]

[Insert Figure 3.1 about here]

3.2 Kernel-based weights

Besides exhaustive grid search, an alternative approach is to use kernel function to generate weights. The kernel method is well-established statistical practice due its simplicity and interpretability. See also Dong and Zheng (2001) for justifications of this

approach. It should be noted that for all cases of r, the set of two-tuples $\{(i, w_i^*)\}_{i=1}^r$ reported in Table 3.2 take symmetric shape and can be very well fitted by degree-two polynomials, with coefficients of determination larger than 0.99 from ordinary least square fit. Therefore, we suggest using the Epanecknikov kernel function $K(u) = 0.75(1-u^2)I_{[-1,1]}(u)$, a symmetric degree-two probability density function, to generate weights. Our proposed kernel-based weight $w_i = g_a^K(i)$ is defined by:

$$g_a^K(i) = K(\frac{2ai + a - ar}{r}) / \sum_{i=1}^{r} K(\frac{2ai + a - ar}{r}),\ i \in D, \tag{3.1}$$

where $a \in [0,1]$ and the denominator normalize f to integrate to 1. Note here that we use kernel evaluation (3.1) with kernel function K truncated at $-a$ and a to generate weights, whereas Dong and Zheng (2001) use subinterval integration of K such that their GEFP integrates to 1 as well. It is straightforward to obtain the AMISE formula (2.3) of $\hat{f}_r$ by using (2.3) with $w_i = g_a^K(i)$. Let g_1^K be the weight function g_a^K with $a = 1$ and g^U the discrete uniform weight function, i.e. $w_i = g^U(i) = 1/r$. It can be shown that g_a^K has the following equivalent form:

$$g_a^K(i) = (1-p)g_1^K(i) + p\frac{1}{r},\ i \in D,$$

where p is a function of a and r defined as:

$$p = \begin{cases} r(a^{-2}-1) \times [r(a^{-2}-1) + \sum_{i=1}^{r} K((2i+1-r)/r)]^{-1} & a \in (0,1], \\ 1 & a = 0, \end{cases}$$

Hence g_a^K can also be expressed by the mixture of g_1 and the discrete uniform weight function g^U. For any fixed value of r, one can choose the optimal value of p (denoted by p^*), such that $w_i = (1-p^*)g_1^K(i) + p^*/r$ lead to minimum value of $C(r)$. Let g_p^* denote the optimal mixture weight function, i.e. $g_p^*(i) = (1-p^*)g_1^K(i) + p^*/r$, and let $\hat{f}_r^K$ and $\hat{f}_r^M$ denote the frequency polygon $\hat{f}_r$ using

the weights defined by g_1^K and g_p^* respectively. Table 3.2 reports the values of $w_i = g_p^*(i)$ (column 4) along with the minimum $C(r)$ values of $\hat{f}_r^*$ and $\hat{f}_r^M$, for $r = 2, 3, 4, 5, 6$ and 7. Observe that g_p^* and g^* have negligible differences and thus $\hat{f}_r^M$ and $\hat{f}_r^*$ have almost the same $C(r)$ values. Those values imply the relationship $\text{AIMSE}^*(\hat{f}_r^M) = a_r \cdot \text{AIMSE}^*(\hat{f}_r^*)$, where for $r = 3, 4, 5, 6$ and 7, $a_r = 1, 1, 1.00001,$ 1.00001 and 1.00001, respectively. Figure 3.1 confirms this result. The dashed and solid lines show the minimum $C(r)$ values respectively, of $\hat{f}_r^K$ and $\hat{f}_r^M$, $r \geq 2$, whereas the six plus sings plot those of $\hat{f}_r^*$, $r = 2, 3, 4, 5, 6$ and 7. It is shown in Figure 3.1 that $\hat{f}_r^M$ based on mixture weight g_p^* has better performance than $\hat{f}_r^K$ and its AMISE performance close resembles that of $\hat{f}_r^*$. Hence, the optimal weights and optimal performances of $\hat{f}_r^*$ are very well approximated by our proposed kernel method for the above values of r.

3.3 Minimum variance weight

Section 3.2 shows that $\hat{f}_r$ based on the mixture of g_1^K and g^U has smaller AMISE than that based on g_1^K. Given any weight function g supported on D, the following proposition further examines the effect of using the mixture weight $m_p = (1-p)g + p/r,\ p \in [0,1]$, through the AIV of $\hat{f}_r$. For this purpose, we impose the following two mild assumptions on g, in addition to the ones on g and f in Section 2.

(A1) g is a symmetric weight function defined on D

(A2) g is nondecreasing on $\{1, 2, \cdots, [r/2]+1\}$

Proposition 3.1. Given a nonnegative weight function g that satisfies (A1) and (A2), define $m_p = (1-p)g + p/r,\ p \in [0,1]$, as the mixture of g and g^U. Let K_{r,m_p} be

defined as (2.2) with $w_i = m_p(i)$. Then $(2/3)\, R(K_{r,m_p}) + (1/3)\, R^*(K_{r,m_p})$, the AIV of $\hat{f}_r$ using m_p, is nonincreasing in p.

Remark 3.1 {Minimum variance weights} Consider $m_p = (1-p)g + p/r,\ p \in [0,1]$, where g is any given weight function that satisfies the assumptions in Proposition 3.1. Let m_0 and m_1 denote m_p using $p = 0$ and $p = 1$, respectively (i.e. $m_0 = g$, $m_1 = g^U$), and let $\hat{f}_r^U$ denote the frequency polygon $\hat{f}_r$ based on g^U. Proposition (3.1) implies that, for m_p with $p \in [0,1]$, $2R(K_{r,m_0})/3 + R^*(K_{r,m_0})/3 \geq 2R(K_{r,m_p})/3 + R^*(K_{r,m_p})/3 \geq 2R(K_{r,m_1})/3 + R^*(K_{r,m_1})/3$. Therefore, for any fixed value of r, the AIV of $\hat{f}_r$ based on m_p is smaller than (or equal to) that of $\hat{f}_r$ using g. Moreover, $\hat{f}_r$ has the smallest AIV when $g = g^U$, among the collection of $\hat{f}_r$ based on the weight functions that satisfy the assumptions in Proposition 3.1. Note that the assumption (A1) and (A2) can't be eliminated from Proposition 3.1. As a counterexample, for $r = 3$, let $g^U(i) = 1/3$, $i = 1, 2$ and 3, and let $g(1) = g(3) = 2/5$, $g(2) = 1/5$. When $\hat{f}_r$ uses g, it follows that $\mathrm{AIV}(\hat{f}_r^U) = (8/27)/(3nb) > \mathrm{AIV}(\hat{f}_r) = (22/75)/(3nb)$. Hence, the discrete uniform weight function doest not lead to minimum variance when assumption (A2) fails to hold. A Similar argument can be drawn for assumption (A1).

Remark 3.2 {Minimum AIMSE^* of $\hat{f}_r^U$} Dong and Zheng (2001) has shown that for even values of r, $\hat{f}_r^U$ has minimum AIMSE^* when $r = 6$. Figure 3.2 plots the $C(r)$ values of $\hat{f}_r^U$ for odd and even values of r, which confirms the result of Dong and Zheng (2001) that AIMSE^* is minimized at $r = 6$ for even value of r (with $C(6) = 0.361334$) and shows that the AIMSE^* is minimized at $r = 5$ (with $C(5) = 0.361262$) for all values of r.

[Insert Figure 3.2 about here]

4. Monte Carlo Simulations

In this section, a Monte Carlo simulation for some small and moderate values of r is carried out to gain some insight into the generalized version of frequency polygon in finite sample situations. The density functions considered for simulations are the following eight normal mixture density functions: Standard Normal, Kurtotic unimodal, Outlier, Bimodal, Separate bimodal, Skewed bimodal, Trimodal and Asymmetric claw (Marron and Wand (1992) and Dong and Zheng (2001)), see the Appendix for their definitions. We shall compare the performance of $\hat{f}_r^*$, $\hat{f}_r^M$ and $\hat{f}_r^U$, for $r = 2$, 3, 4, 5, 6 and 7. To run the simulations, we generate random sample from each density for four different sample sizes $n = 30$, 50, 100, 200. For each frequency polygon and a given bin width, we use the anchor placement rule $\min\{X_1, X_2, \cdots, X_n\} - b/10$ (see Jones et al. (1998) and Dong and Zheng (2001)) to determine the placement of the bins. The accuracies of the above three estimators are assessed by their sample MISE values. For this purpose, we first generate 1000 repetitions of random samples from each density for each sample size. Then for $\hat{f}_r^*$, we numerically approximate each sample's respective integrated square error $\mathrm{ISE}(\hat{f}_r^*) = \int \left(\hat{f}_r^*(x) - f(x) \right)^2 dx$ by $\mathrm{ISE}(\hat{f}_r^*) = (1/400) \times \sum_{i=1}^{401} [\hat{f}_r^*(t_i) - f(t_i)]^2$, where $t_i = -4 + (i-1)(8/400)$, $i = 1, \cdots, 401$. The formulae for $\hat{f}_r^M$ and $\hat{f}_r^U$ are defined similarly. The simulated versions of these estimators' MISE are thus approximated by the average of the 1000 sample's ISE values. Given the value of r and sample size n, the MISE value is determined by the value of bin width b. For each set of 1000 samples, we use 1000 equally spaced grid values of bin width to computed their associated MISE values and

obtain the minimal value of the simulated MISE.

We use the minimal quantities of the simulated MISE's to compare the MISE ratios $\text{MISE}(\hat{f}_r^*)/\text{MISE}(\hat{f}_2)$, $r = 3,\ 4,\ 5,\ 6$ and 7. The ratios for $\hat{f}_r^M$ and $\hat{f}_r^U$ are computed analogously. As demonstrated by the simulation results, the simulated MISE ratios associated with $\hat{f}_r^M$ and $\hat{f}_r^*$ round to three decimal places are the same for the bulk of all the ratios obtained from simulations, no doubt thanks to the closeness of the two weights g^* and g_p^* reported in Table 3.2. So, we report only the MISE ratios associated with $\hat{f}_r^M$ and $\hat{f}_r^U$. Table 4.1 reports the ratio $\text{MISE}(\hat{f}_r^M)/\text{MISE}(\hat{f}_2)$ and $\text{MISE}(\hat{f}_r^U)/\text{MISE}(\hat{f}_2)$. The latter ratios are reported in the parenthesis. The MISE ratios smaller than 1 indicate that the simulated minimal MISE's of $\hat{f}_r^M$ are smaller than that of EFP. Since the ratios for $\hat{f}_r^M$ in Tables 4.1 is smaller than 1 for all sample sizes and all values of r, the frequency polygon $\hat{f}_r^M$, $r = 3, 4, 5, 6$ and 7, improve the accuracy of EFP for the above eight densities and the accuracy improves as r increases. The above results show that the improvement of $\hat{f}_r^M$ in term of AMISE demonstrated in Table 3.2 carries over to the situations of small and moderate sample size. Observe also that the MISE ratios associated with $\hat{f}_r^U$ in the parenthesis in all entries are larger than those associated with $\hat{f}_r^M$ and most of these ratios in the parenthesis are minimized when r is 4, 5 or 6. So, $\hat{f}_r^M$ has better performance than $\hat{f}_r^U$ for the above eight densities in finite sample situations and the MISE ratios associated with $\hat{f}_r^U$ in general concur with the conclusion drawn in Remark 3.2.

[Insert Table 4.1 about here]

Appendix

Proof of Proposition 3.1.

We now give the proof of Proposition 3.1 only for even value of r. The proof for the case of odd value of r is similar and hence is omitted.

Since $2R(K^*_{r,g})/r = \sum_{i=1}^{r}(g(i))^2$, and by the symmetry of g, one has:

$$\begin{aligned}2R^*(K^*_{r,g})/r &= \sum_{i=1}^{r-1} g(i)g(i+1)\\ &=[\sum_{i=1}^{r-1}(g(i))^2 + \sum_{i=1}^{r-1}(g(i+1))^2 - \sum_{i=1}^{r-1}(g(i+1)-g(i))^2]/2\\ &= \sum_{i=1}^{r}(g(i))^2 - (g(r))^2 - \sum_{i=1}^{r/2-1}(g(i+1)-g(i))^2\end{aligned}$$

Consequently,

$$4R(K^*_{r,g})/3r + 2R^*(K^*_{r,g})/3r = \sum_{i=1}^{r}(g(i))^2 - \sum_{i=1}^{r/2-1}(g(i+1)-g(i))^2/3 - (g(r))^2/3.$$

By the definitions of $R(K^*_{r,m_p})$, $R^*(K^*_{r,m_p})$ and m_p, the last formula can be derived for:

$$\begin{aligned}&4R(K^*_{r,m_p})/3r + 2R^*(K^*_{r,m_p})/3r\\ &= \sum_{i=1}^{r}((1-p)g(i)+p/r)^2\\ &\quad - \sum_{i=1}^{r/2-1}[(1-p)g(i+1)+p/r-((1-p)g(i)+p/r)]^2/3\\ &\quad - ((1-p)g(r)+p/r)^2/3\\ &= 1/r + (1-p)^2\sum_{i=1}^{r}(g(i)-1/r)^2\\ &\quad - (1-p)^2\sum_{i=1}^{r/2-1}(g(i)-g(i+1))^2/3\\ &\quad - (1-p)^2(g(r))^2/3 - 2(1-p)pg(r)/3r - p^2/3r^2.\end{aligned}$$

Differentiating the right hand side with respect to p yields:

$$\begin{aligned}&(d/dp)\left(4R(K^*_{r,m_p})/3r + 2R^*(K^*_{r,m_p})/3r\right)\\ &= -2(1-p)\left[\sum_{i=1}^{r}(g(i)-1/r)^2 - \sum_{i=1}^{r/2-1}(g(i)-g(i+1))^2/3\right]\\ &\quad + 2(1-p)g(r)(g(r)-1/r)/3 + 2p(g(r)-1/r)/3r \le 0,\end{aligned}$$

The last two terms on the right hand side are obviously nonnegative. The proof the Proposition 3.1 is complete by showing that the first term is also nonnegative. To see

this, recall that g is nondecreasing and symmetric, there exists an integer j_r, $1 < j_r < r/2$, such that $g(j_r) \le 1/r \le g(j_r+1)$. Hence, $(g(i)-1/r)^2 \ge (g(i)-g(i+1))^2$, for $i \le j_r - 1$ or $i > j_r$. In addition, by the symmetry of g, $g(j_r+1) = g(r-j_r)$. Therefore,

$$
\begin{aligned}
&\textstyle\sum_{i=1}^{r}(g(i)-1/r)^2 - \sum_{i=1}^{r/2-1}(g(i)-g(i+1))^2/3 \\
&\ge \textstyle\sum_{i=1}^{r/2-1}(g(i)-1/r)^2 - \sum_{i=1}^{r/2-1}(g(i)-g(i+1))^2/3 + (g(r-j_r)-1/r)^2 \\
&\ge \textstyle\sum_{i=1}^{j_r-1}(g(i)-1/r)^2 + (g(j_r)-1/r)^2 + \sum_{i=j_r+1}^{r/2-1}(g(i)-1/r)^2 + (g(r-j_r)-1/r)^2 \\
&\quad - \Big[\textstyle\sum_{i=1}^{j_r-1}(g(i)-g(i+1))^2/3 + (g(j_r)-g(j_r+1))^2/3 + \sum_{i=j_r+1}^{r/2-1}(g(i)-g(i+1))^2/3\Big] \\
&\ge (g(j_r)-1/r)^2 + (g(r-j_r)-1/r)^2 - \big[g(j_r)-1/r+1/r-g(j_r+1)\big]^2/3 \\
&= 2\,(g(j_r)-1/r)^2/3 + 2(1/r-g(j_r+1))^2/3 - 2(g(j_r)-1/r)(1/r-g(j_r+1))/3 \\
&\ge 2\,(g(j_r)-1/r)^2/3 + 2(1/r-g(j_r+1))^2/3 - \max\{2(g(j_r)-1/r)^2/3,\, 2(1/r-g(j_r+1))^2/3\} \\
&= \min\{2(g(j_r)-1/r)^2/3,\, 2\,(1/r-g(j_r+1))^2/3\} \ge 0
\end{aligned}
$$

Normal mixtures used in the simulation:

Standard normal: $N(0,1)$.

Kurtotic unimodal: $(2/3)\,N(0,1) + (1/3)\,N(0, 0.1^2)$.

Outlier: $0.1\,N(0,1) + 0.9\,N(0, 0.1^2)$.

Bimodal: $0.5\,N(-1, (2/3)^2) + 0.5\,N(1, (2/3)^2)$.

Separated bimodal: $0.5\,N(-1.5, 0.5^2) + 0.5\,N(1.5, 0.5^2)$.

Skewed bimodal: $0.75\,N(0, 1) + 0.25\,N(1.5, (1/3)^2)$.

Trimodal: $0.45\,N(-1.2, 0.6^2) + 0.45\,N(1.2, 0.6^2) + 0.1\,N(0, 0.25^2)$.

Asymmetric claw: $0.5\,N(0, 1) + \sum_{\ell=-2}^{2}(2^{\ell-1}/31)N(\ell + 0.5,\, (2^{-\ell}/10)^2)$.

Reference

Dong, J.P. and Zheng, C. (2001). Generalized edge frequency polygon for density estimation. Statistics and Probability Letters, 55, 137-145.

Jones, M. C. (1989). Discretized and interpolated kernel density estimates. Journal of the American Statistical Association, 84, 733-741.

Jones, M. C., Samiuddin M., Al-Harbey, A.H. and Maatouk, T.A.H. (1998). The edge frequency polygon. Biometrika, 85, 235-239.

Scott, D. W. (1985a). Frequency polygons: theory and application. Journal of the American Statistical Association, 80, 348-354.

Scott, D. W. (1985b). Average shifted histograms: effective nonparametric density estimations in several dimensions. Annals of Statistics, 13, 1024-1040.

Scott, D.W. (1992). Multivariate Density Estimation: Theory, Practice, and Visualization. John, Wiley & Sons.

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

Simonoff, J.S. and Unida, F. (1997) Measuring the stability of histogram appearance when the anchor position is changed. Computational Statistics and Data Analysis, 23, 335-353.

Marron, J.S. and Wand, M.P. (1992). Exact mean integrated squared error. Annals of Statististics, 20, 712 – 736.

Table 3.1 List of notations concerning $\hat{f}_r$

$\hat{f}_r^*$: $\hat{f}_r$ using optimal weight function g^* reported in Table 3.2, (Section 3.1).
$\hat{f}_r^K$: $\hat{f}_r$ using g_1^K defined as (3.1) with $a=1$, (Section 3.2).
$\hat{f}_r^M$: $\hat{f}_r$ using optimal mixture weight $g_p^* = (1-p^*)g_1^K + p^*/r$, , for a given g, (Section 3.2).
$\hat{f}_r^U$: $\hat{f}_r$ using discrete uniform weight g^U, $g^U(i) = 1/r, i \in D$, (Section 3.3).

Table 3.2

r	i $g^*(i)=w_i^*$	$C(r)$ of $\hat{f}_r^*$	i $g_p^*(i)$	$C(r)$ of $\hat{f}_r^M$
3	1 0.31024169 2 0.37951661 3 0.31024169	0.36435138	1 0.31024169 2 0.37951661 3 0.31024169	0.36435138
4	1 0.20898349 2 0.29101651 3 0.29101651 4 0.20898349	0.35863201	1 0.20898349 2 0.29101651 3 0.29101651 4 0.20898349	0.35863201
5	1 0.15073325 2 0.22352424 3 0.25148502 4 0.22352424 5 0.15073325	0.35563429	1 0.15013011 2 0.22493494 3 0.24986989 4 0.22493494 5 0.15013011	0.35563811
6	1 0.11375116 2 0.17608115 3 0.21016769 4 0.21016769 5 0.17608115 6 0.11375116	0.35386467	1 0.11301611 2 0.17739678 3 0.20958711 4 0.20958711 5 0.17739678 6 0.11301611	0.35386843
7	1 0.08885989 2 0.14170879 3 0.17604945 4 0.18676374 5 0.17604945 6 0.14170879 7 0.08885989	0.35273001	1 0.088132659 2 0.14285714 3 0.17569183 4 0.18663673 5 0.17569183 6 0.14285714 7 0.088132659	0.35273330

Table 4.1

samlpe size	n=30	n=50	n=100	n=200
r		Gaussian		
3	0.973 (0.972)	0.956 (0.960)	0.948 (0.955)	0.956 (0.960)
4	0.948 (0.960)	0.941 (0.952)	0.936 (0.954)	0.937 (0.955)
5	0.940 (0.956)	0.924 (0.950)	0.918 (0.950)	0.926 (0.950)
6	0.937 (0.963)	0.923 (0.954)	0.9183 (0.949)	0.920 (0.953)
7	0.925 (0.968)	0.916 (0.955)	0.913 (0.954)	0.917 (0.956)
		Kurtotic		
3	0.982 (0.998)	0.977 (0.988)	0.970 (0.980)	0.956 (0.960)
4	0.974 (1.002)	0.966 (0.987)	0.958 (0.978)	0.937 (0.955)
5	0.968 (1.011)	0.959 (0.992)	0.952 (0.979)	0.926 (0.950)
6	0.965 (1.015)	0.956 (0.998)	0.949 (0.983)	0.920 (0.953)
7	0.962 (1.020)	0.954 (1.0006)	0.946 (0.986)	0.917 (0.956)
		Outlier		
3	0.965 (0.973)	0.955 (0.960)	0.962 (0.964)	0.962 (0.966)
4	0.944 (0.963)	0.935 (0.954)	0.948 (0.961)	0.945 (0.961)
5	0.933 (0.970)	0.929 (0.958)	0.937 (0.957)	0.937 (0.957)
6	0.931 (0.968)	0.920 (0.960)	0.930 (0.965)	0.931 (0.958)
7	0.928 (0.970)	0.921 (0.962)	0.930 (0.965)	0.928 (0.963)
		Bimodal		
3	0.996 (0.982)	0.973(0.980)	0.963 (0.973)	0.965 (0.970)
4	0.997 (0.967)	0.956 (0.981)	0.941 (0.968)	0.948 (0.967)
5	0.959 (1.002)	0.951(0.985)	0.931 (0.962)	0.940 (0.963)
6	0.959 (1.014)	0.945(0.990)	0.932 (0.965)	0.936 (0.965)
7	0.958 (1.017)	0.943(0.991)	0.923 (0.971)	0.932 (0.969)
		Separate bimodal		
3	0.952 (0.961)	0.962 (0.969)	0.972 (0.975)	0.961 (0.966)
4	0.949 (0.968)	0.950 (0.967)	0.953 (0.970)	0.946 (0.960)
5	0.939 (0.969)	0.940 (0.965)	0.940 (0.967)	0.938 (0.960)
6	0.934 (0.973)	0.932 (0.970)	0.937 (0.970)	0.932 (0.961)
7	0.932 (0.976)	0.932 (0.973)	0.934 (0.970)	0.930 (0.962)
		Skewed bimodal		
3	0.980 (0.995)	0.975 (0.991)	0.973 (0.981)	0.969 (0.975)
4	0.968 (1.007)	0.964 (0.997)	0.953 (0.979)	0.957 (0.976)
5	0.959 (1.015)	0.958 (0.999)	0.946 (0.981)	0.949 (0.971)
6	0.959 (1.021)	0.950 (1.005)	0.942 (0.984)	0.945 (0.978)
7	0.956 (1.025)	0.949 (1.011)	0.940 (0.988)	0.942 (0.979)

Table 4.1 (continued)

samlpe size	n=30	n=50	n=100	n=200
r		Trimodal		
3	0.981 (0.998)	0.971 (0.988)	0.972 (0.985)	0.976 (0.984)
4	0.971 (1.006)	0.966 (1.000)	0.959 (0.986)	0.963 (0.985)
5	0.968 (1.021)	0.962 (1.008)	0.954 (0.987)	0.955 (0.986)
6	0.968 (1.030)	0.961 (1.013)	0.947 (0.995)	0.952 (0.991)
7	0.965 (1.035)	0.956 (1.018)	0.945 (1.000)	0.949 (0.993)
		Asymmetric claw		
3	0.987 (1.002)	0.987 (1.002)	0.989 (1.006)	0.989 (1.005)
4	0.984 (1.006)	0.984 (1.006)	0.983 (1.018)	0.979 (1.013)
5	0.974 (0.996)	0.974 (0.996)	0.977 (1.026)	0.975 (1.023)
6	0.971 (0.992)	0.971 (0.992)	0.974 (1.030)	0.972 (1.029)
7	0.968 (0.990)	0.968 (0.989)	0.972 (1.036)	0.971 (1.034)

The ratios $\mathrm{MISE}(\hat{f}_r^M)/\mathrm{MISE}(\hat{f}_2)$ and $\mathrm{MISE}(\hat{f}_r^U)/\mathrm{MISE}(\hat{f}_2)$ (in parenthesis), for $r = 3, 4, 5, 6, 7$

CAPTIONS

Figure 3.1. Plus signs are the values of $C(r)$ of $\hat{f}_r^*$, for $r = 2, 3, 4, 5, 6$ and 7. Dashed and solid curves plot those values of $\hat{f}_r^K$ and $\hat{f}_r^M$, $r \geq 2$, respectively. Note that $\hat{f}_2^*$, $\hat{f}_2^M$, $\hat{f}_2^K$ and $\hat{f}_2^U$ are the same, since they all use $w_1 = w_2 = 0.5$ by assumption.

Figure 3.2. The values of $C(r)$ of $\hat{f}_r^U$ plotted against the value of r, with minimum value at $r = 5$.

Figure 3.1

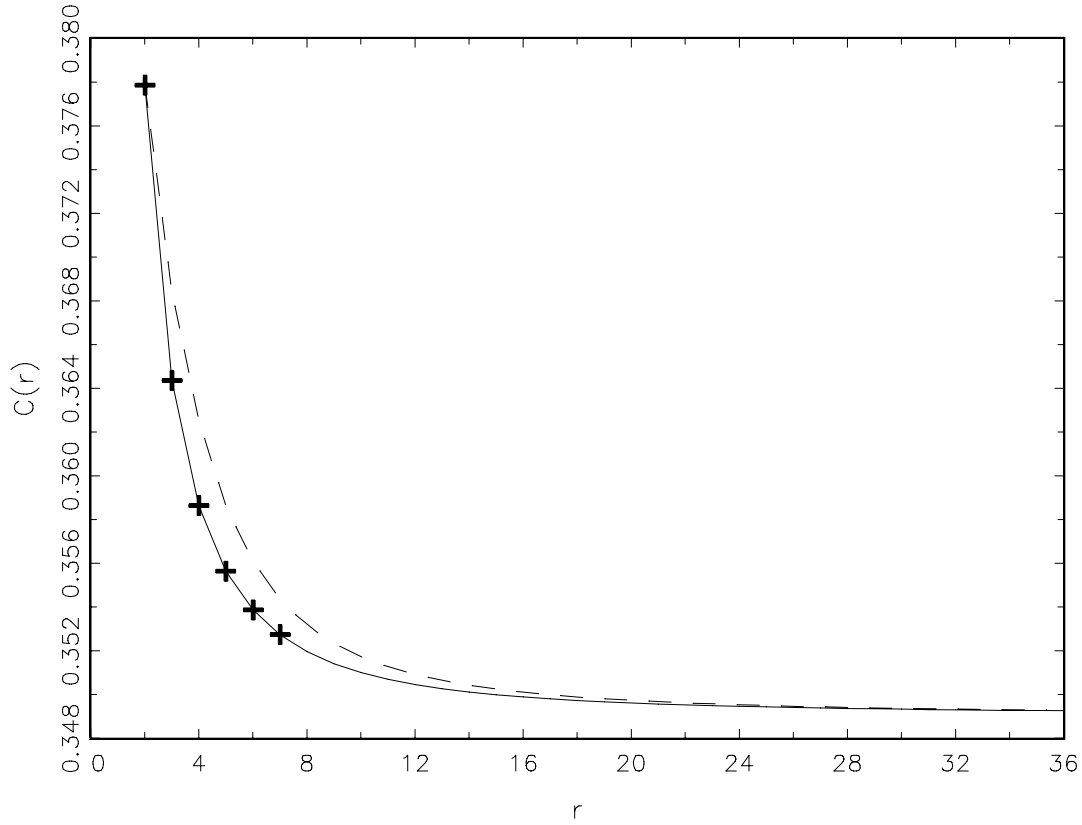
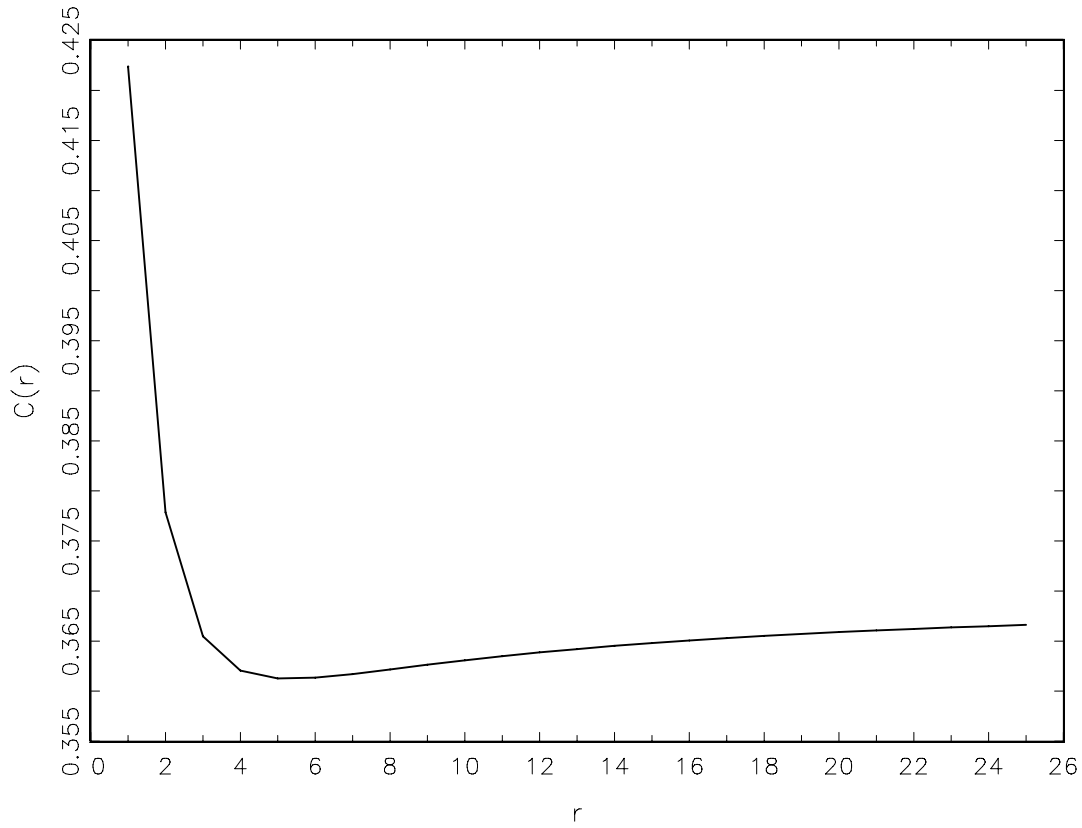

C(r)
0.348
0.352
0.356
0.360
0.364
0.368
0.372
0.376
0.380
0
4
8
12
16
20
24
28
32
36
r

Figure 3.2


C(r)
0.355
0.365
0.375
0.385
0.395
0.405
0.415
0.425
0
2
4
6
8
10
12
14
16
18
20
22
24
26
r