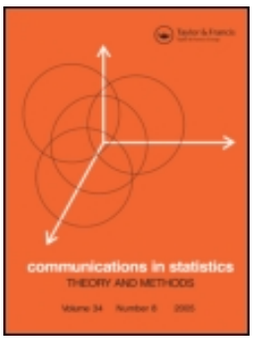

Taylor & Francis
Taylor & Francis Group

Communications in Statistics - Theory and Methods

ISSN: 0361-0926 (Print) 1532-415X (Online) Journal homepage: http://www.tandfonline.com/loi/lsta20

A Note on the Frequency Polygon Based on the Weighted Sums of Binned Data

Wen-shuenn Deng, Jyh-shyang Wu, Li-ching Chen & Shun-jie Ke

To cite this article: Wen-shuenn Deng, Jyh-shyang Wu, Li-ching Chen & Shun-jie Ke (2014) A Note on the Frequency Polygon Based on the Weighted Sums of Binned Data, Communications in Statistics - Theory and Methods, 43:8, 1666-1685, DOI: 10.1080/03610926.2012.673675

To link to this article: https://doi.org/10.1080/03610926.2012.673675

Communications in Statistics—Theory and Methods, 43: 1666–1685, 2014

ISSN: 0361-0926 print / 1532-415X online
DOI: 10.1080/03610926.2012.673675

Taylor & Francis
Taylor & Francis Group

A Note on the Frequency Polygon Based on the Weighted Sums of Binned Data

WEN-SHUENN DENG,[1] JYH-SHYANG WU,[2] LI-CHING CHEN,[1] AND SHUN-JIE KE[2]

[1]Department of Statistics, Tamkang University, New Taipei City, Taiwan
[2]Department of Mathematics, Tamkang University, New Taipei City, Taiwan

We revisit the generalized midpoint frequency polygons of Scott (1985), and the edge frequency polygons of Jones et al. (1998) and Dong and Zheng (2001). Their estimators are linear interpolants of the appropriate values above the bin centers or edges, those values being weighted averages of the heights of r, $r \in N$, *neighboring histogram bins. We propose a simple kernel evaluation method to generate weights for binned values. The proposed kernel method can provide near-optimal weights in the sense of minimizing asymptotic mean integrated square error. In addition, we prove that the discrete uniform weights minimize the variance of the generalized frequency polygon under some mild conditions. Analogous results are obtained for the generalized frequency polygon based on linearly prebinned data. Finally, we use two examples and a simulation study to compare the generalized midpoint and edge frequency polygons.*

Keywords Edge frequency polygon; Kernel-based weights; Midpoint frequency polygon; Minimum variance weights; Mixture weights; Uniform weights.

Mathematics Subject Classification Primary 62G07; Secondary 62G20.

1. Introduction

Given a univariate random sample of size n from an unknown continuous distribution, consider the problem of estimating the distribution's density function f via variants of histogram (binned) values. The frequency polygon, e.g., Scott (1985), is a simple density estimator based on histogram, with some form of linear interpolation. Although researchers have proposed many different density estimators, frequency polygon remains a widely used density estimator, no doubt thanks to the computational simplicity of its constituent histogram values and the statistical efficiency comparable with nonnegative kernel density estimators; see Scott (1985a, 1992) for excellent reviews of such density estimator. See also Jones (1989) and Lin et al. (2006) for linearly interpolated density based on kernel density estimates. When triangular kernel function is employed to obtain kernel estimates, their estimators can be viewed as linear interpolants of linearly prebinned data.

Received August 19, 2011; Accepted March 1, 2012.
Address correspondence to Wen-Shuenn Deng, Department of Statistics, Tamkang University, 151 Ying-Chuang Road, Tamsui, New Taipei City, Taiwan. E-mail: 121350@ mail.tku.edu.tw

A simple approach to improve the efficiency of the frequency polygon without spoiling its simplicity is to interpolate the weighted sums of a collection of histogram values above bin centers or edges. Let r be the number of histogram values used to obtain such weighted sums, then the uses of odd and even values of r result in midpoint and edge frequency polygons, respectively. In this article, we consider density estimator defined as the linear interpolant of such weighted sums of binned values and investigate the choices of weights. Call the above two estimators simply frequency polygon if there is no danger of confusion in what follows. As we will see in the next section, when r is even, the studied frequency polygon corresponds to the edge frequency polygon (EFP) proposed by Jones et al. (1998) and its generalized version known as generalized edge frequency polygon (GEFP) considered by Dong and Zheng (2001). For odd values of r, it corresponds to the frequency polygon of averaged shifted histograms (FP-ASH) of Scott (1992, 2010).

Given a bin width, it is the choice of weights that exclusively determine the asymptotic mean integrated square error (AMISE) of the frequency polygon. The main purpose of this article is to propose a simple kernel evaluation method to generate weights for binned values. In Sec. 3, we prove that our proposed kernel method can provide near optimal weights with respect to its AMISE. As a continuation of Dong and Zheng (2001), we obtain, in addition, the optimal weights for $r = 3, 5, 7$ and show that the minimal AMISE monotonically decreases as the value of r increases from 2 to 7 and that the frequency polygons have similar efficiencies when r exceeds 5.

The simplest weight is the uniform weight. In Sec. 3, we further show that the use of uniform weight leads to the minimum variance. It is shown in Dong and Zheng (2001) that the GEFP based on uniform weight has minimum AMISE at $r = 6$. In this article, we further show that the generalized frequency polygon for all values of r has minimum AMISE when $r = 5$.

An alternative rule to bin the data is the linear binning rule. The frequency polygon based on linearly prebinned data also inherently integrates to 1, as is the case of the histogram based frequency polygon. In Sec. 4, we prove that our proposed kernel method also provides near optimal weight when frequency polygon uses weighted sum of linearly prebinned data. As in the case of histogram-based frequency polygon, we also prove that the uniform weight also leads to the minimum variance of frequency polygon using linearly prebinned data.

In Sec. 5, the generalized midpoint frequency polygon is found to be better at picking up the bumps of densities than the generalized edge frequency polygon. We explained the rationale in conjunction with a real data example and a simulated data example. A simulation study is carried out to compare the efficiencies of the generalized midpoint and edge frequency polygons in finite sample situations. The simulation results concur with the findings in Sec. 3.

2. Generalized Midpoint and Edge Frequency Polygons

Let $X_1, X_2, \ldots, X_n$ 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 $\sum_{k \in Z} n_k = n$ by assumption.

We now give the definition of frequency polygon as follows. Here and throughout this article, let g be a non negative and symmetric weight function defined on $D = \{1, 2, \ldots, r\}$

such that $\sum_{i=1}^{r} g(i) = 1$. Given a weight function g, let $w_i = g(i)$, $i = 1, 2, \ldots, r$, be the weights used to obtain weighted sum of binned values. Define the weighted average of r neighboring histogram (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 defined as:

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

for odd and even values of r, respectively.

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 EFP considered by Jones et al. (1998). The latter authors point out that 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) demonstrated that the appearance of $\hat{f}_{\mathrm{EFP},2}$ is less sensitive to the choice of anchor position (bin origin). On the other hand, when r is odd, $\hat{f}_{\mathrm{MFP},r}$ is the generalized version of midpoint frequency polygon. Scott (1985b, 2012) average over q histograms of different anchors but a common bin width, say h, 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 weight function $w_j = g(i) = 1 - |j - q|/q$, $j = 1, 2, \ldots, 2q - 1$, in terms of our notations. FP-ASH and GEFP are computationally and statistically efficient density estimators. In addition, they can alleviate the noise effect of anchor; see Scott (1985b, 1992, 2010) for detailed discussions on FP-ASH. See also Simonoff and Udina (1997) for their insight into the stability of EFP's appearance when anchor position is changed. From now on, we 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 approaches of Scott (1985b), Jones et al. (1998), and Dong and Zheng (2001), we will use the AMISE of $\hat{f}_r$, $r \geq 2$, to understand the effect of the above generalization on frequency polygon. The result (formula 3.2) of Jones (1989) can serve this purpose if u_k and v_k are rewritten in form of kernel estimates. To this end, we reformulate u_k and v_k as follows. Let

$$K_r(x) = \sum_{j=1}^{r} \frac{r}{2} w_j I_{((2j-r-2)/r,(2j-r)/r]}(x). \tag{2}$$

It can be shown that K_r is a symmetric probability density function supported on $[-1, 1]$, and that u_k and v_k can be rewritten in form of kernel estimates using bin width $br/2$ as

$$u_k = \frac{1}{nbr/2} \sum_{i=1}^{n} K_r\left(\frac{X_i - t_k}{br/2}\right),$$

$$v_k = \frac{1}{nbr/2} \sum_{i=1}^{n} K_r\left(\frac{X_i - s_k}{br/2}\right).$$

The following proposition follows directly from formula (3.2) of Jones (1989).

Proposition 2.1. *Suppose that the density* f *has continuous second derivative* f'' *with* $R(f'') = \int f''(x)^2 dx < \infty$ *and that* f *is ultimately monotone. Let* g *be a non negative symmetric weight function defined on* D. *If* $n \to \infty, b \to 0$ *and* $n\,b \to \infty$, *then the mean integrated square error (MISE) of* $\hat{f}_r$, $MISE(\hat{f}_r) = E[\int\{\hat{f}_r(x) - f(x)\}^2 dx]$, $r \geq 1$, *is given by:*

$$MISE(\hat{f}_r) = \frac{R(f'')b^4}{4}\left[\mu_2^2 + \frac{\mu_2}{3} + \frac{1}{30}\right] + \frac{1}{nb}\left[\frac{2}{3}R(K_r) + \frac{1}{3}R^*(K_r)\right] + o\left(\frac{1}{nb} + b^4\right), \tag{3}$$

where $\mu_2 = (r/2)^2 \int x^2 K_r(x)dx = \sum_{i=1}^{r} g(i)[(i - (r+1)/2)^2 + 1/12]$, $R(K_r) = (r/2)^{-1} \times \int K_r(x)^2 dx = \sum_{i=1}^{r} g(i)^2$ *and* $R^*(K_r) = (r/2)^{-1} \int K_r(x)K_r(x - 2/r)dx = \sum_{i=1}^{r-1} g(i)\, g(i+1)$, *for* $r \geq 2$ *and* $R^*(K_r) = 0$, *for* $r = 1$.

The leading terms of MISE($\hat{f}_r$) are, respectively, the asymptotic integrated square bias (AISB) and asymptotic integrated variance (AIV). Let their sum be denoted as AMISE($\hat{f}_r$). Given a value of r, one seeks to obtain weights w_i such that AMISE is minimized. We will report the optimal weights w_i for $r = 3, 4, \ldots, 7$, that serve the above purpose. Toward this end, we use the following AMISE-optimal bin width

$$b_r^* = \left(\frac{\frac{2}{3}R(K_r) + \frac{1}{3}R^*(K_r)}{nR(f'')[\mu_2^2 + \frac{1}{3}\mu_2 + \frac{1}{30}]}\right)^{1/5} \tag{4}$$

to obtain the minimum AMISE of $\hat{f}_r$ given by

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

where $C(r) = [\mu_2^2 + \mu_2/3 + 1/30]^{1/5}[2R(K_r)/3 + R^*(K_r)/3]^{4/5}$, $r \geq 1$. Given a value of r, the optimal choice of w_i is thus defined as the minimizer of $C(r)$. For $r = 3, 4, \ldots, 7$, we obtain the optimal weights through exhaustive grid search and report them in Table 1. In the next section, we will propose a simple kernel method to generate weights and prove that the proposed method produces near optimal weights.

3. The Weights

3.1. Kernel-Based Weights

The kernel method is a widely used statistical practice due to its simplicity and interpretability in generating weights; see Dong and Zheng (2001) for justifications of this approach. It can be seen in Table 1 that the optimal weights for each $r \geq 3$ are all positive and mound-shaped. This leads us to approximate the optimal weights by the mixture of a kernel-based weights and the uniform weight. The kernel based weights are generated by the normalized degree-two Epanecknikov kernel function as follows:

$$g^K(i) = K\left(\frac{2i - 1 - r}{r}\right) \Big/ \sum_{i=1}^{r} K\left(\frac{2i - 1 - r}{r}\right), \quad i \in D,$$

Table 1
The optimal weights and $C(r)$ values for $\hat{f}_r^*$ and $\hat{f}_r^M$, $r = 3, 4, \ldots, 7$

r	i	w_i^*	$C(r)$ of $\hat{f}_r^*$	i	$w_i = g_{p^*}(i)$	$C_p(r)$ of $\hat{f}_r^M$	p^*
	1	0.31024169		1	0.31024169		
3	2	0.37951661	0.36435138	2	0.37951661	0.36435138	0.67094
	3	0.31024169		3	0.31024169		
	1	0.20898349		1	0.20898349		
	2	0.29101651		2	0.29101651		
4	3	0.29101651	0.35863201	3	0.29101651	0.35863201	0.54882
	4	0.20898349		4	0.20898349		
	1	0.15073325		1	0.15013011		
	2	0.22352424		2	0.22493494		
5	3	0.25148502	0.35563429	3	0.24986989	0.35563811	0.47014
	4	0.22352424		4	0.22493494		
	5	0.15073325		5	0.15013011		
	1	0.11375116		1	0.11301611		
	2	0.17608115		2	0.17739678		
6	3	0.21016769	0.35386467	3	0.20958711	0.35386843	0.41252
	4	0.21016769		4	0.20958711		
	5	0.17608115		5	0.17739678		
	6	0.11375116		6	0.11301611		
	1	0.08885989		1	0.08813266		
	2	0.14170879		2	0.14285714		
	3	0.17604945		3	0.17569183		
7	4	0.18676374	0.35273001	4	0.18663673	0.35273330	0.36794
	5	0.17604945		5	0.17569183		
	6	0.14170879		6	0.14285714		
	7	0.08885989		7	0.08813266		

where $K(u) = 0.75(1-u^2)I_{[-1,1]}(u)$. Let g^U be the uniform weight, namely, $g^U(i) = 1/r$, $i \in D$. Our proposed weights $w_i = g_p(i)$ is defined as the mixture of g^K and g^U as follows:

$$g_p(i) = (1-p)g^K(i) + p/r, \quad i \in D, \quad p \in [0,1]. \tag{5}$$

For any fixed value of r, now the value of $C(r)$ depends on the choice of p if $\hat{f}_r$ uses $w_i = g_p(i) = (1-p)g^K(i) + p/r$, $i \in D$, We now use $C_p(r)$ to denote the value of $C(r)$ when the weights of $\hat{f}_r$ are generated by g_p. Let p^* denote the optimal choice of p that minimizes $C_p(r)$ and let g_{p^*} be its associated optimal mixture weight function, i.e., $g_{p^*}(i) = (1-p^*)g^K(i) + p^*/r, i \in D$. Let $\hat{f}_r^M$ denote the frequency polygon $\hat{f}_r$ using the weights $w_i = g_{p*}(i)$. Table 1 reports the values of $w_i = g_{p^*}(i)$ (column 6) along with the values of p^* (column 8) and their associated values of $C_p(r)$ (column 7) for $r = 3, 4, \ldots, 7$.

Table 1 also reports the AMISE-optimal weights w_i^* (column 3) among all choices of weights. They are obtained through exhaustive grid search. Let $\hat{f}_r^*$ denote the $\hat{f}_r$ using w_i^*. Observe that $w_i = g_{p*}(i)$ and w_i^* have negligible differences and thus the values of $C_p(r)$ and $C(r)$ of $\hat{f}_r^M$ and $\hat{f}_r^*$ (columns 7 and 4), respectively, have almost the same values. Those values imply the relationship $\text{AMISE}^*(\hat{f}_r^M) = a_r \cdot \text{AMISE}^*(\hat{f}_r^*)$, where for $r = 3, 4, \ldots, 7$,

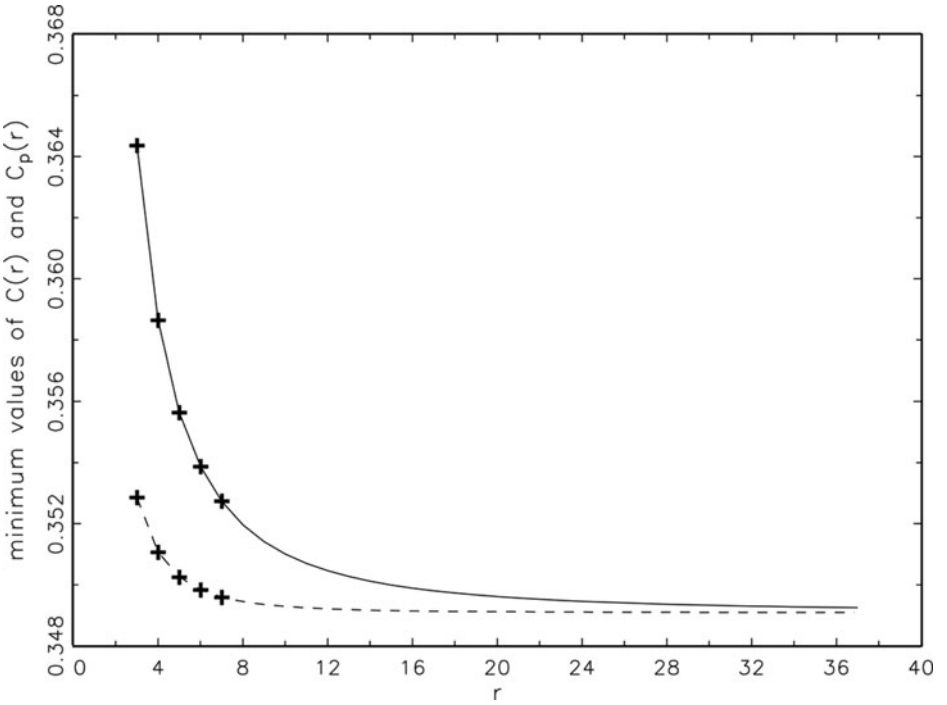

Figure 1. Values of $C(r)$ of $\hat{f}_r^*$ (upper plus signs) and $C_p(r)$ of $\hat{f}_r^M$ (upper solid curve), for $r = 3$, 4, 5, 6, and 7. The lower plus signs and dashed curve plot those of $\tilde{f}_r^*$ and $\tilde{f}_r^M$, respectively.

$a_r = 1, 1, 1.00001, 1.00001$, and 1.00001, respectively. Figure 1 confirms this result. The upper solid curve shows the minimum $C_p(r)$ values of $r = 3, 4, \ldots, 7$, whereas the upper five plus sings plot the minimum $C(r)$ values 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. With brief descriptions, Table 2 helps to clarify some notations used previously and some used in later sections.

Table 2
Summary of frequency polygons and their respective weights

Frequency polygon	Weights used
	simple (histogram) binning rule
$\hat{f}_r$	general weights $w_i = g(i)$
$\hat{f}_r^*$	optimal weights w_i^*
$\hat{f}_r^M$	optimal mixture weights $w_i = (1 - p^*)g^K(i) + p^*/r$
$\hat{f}_r^U$	uniform weights $w_i = g^U(i) = 1/r$
	linear binning rule
$\tilde{f}$	general weights $w_i = g(i)$
$\tilde{f}_r^*$	optimal weights w_i^*
$\tilde{f}_r^M$	optimal mixture weights $w_i = (1 - p^*)g^K(i) + p^*/r$
$\tilde{f}_r^U$	uniform weights $w_i = g^U(i) = 1/r$

3.2. The Minimum Variance Weights

Section 3.1 shows that our proposed simple kernel method can generate near optimal weights. Here, we restrict our attention to the effect of using the mixture of any given weight function g and the uniform weight g^U through the AIV of $\hat{f}_r$. To this end, we need the following the assumptions

(A1) g is a non negative and symmetric weight function defined on $D = \{1, 2, \ldots, r\}$ such that $\sum_{i=1}^{r} g(i) = 1$. (A1) which is introduced in Sec. 2 and (A2).
(A2) g is non decreasing on $\{1, 2, \ldots, [r/2] + 1\}$.

The following proposition suggests that, given an arbitrary weight function g, the AIV of $\hat{f}_r$ based on the mixture weight function $m_p = (1 - p)g + p/r$, $p \in [0, 1]$, decreases monotonically as p increases.

Proposition 3.1. *Given a non negative weight function* g *that satisfies* (A1) *and* (A2), *define the mixture of* g *and* g^U *as* $m_p = (1 - p)g + p/r$, $p \in [0, 1]$. *Let* K_{r,m_p} *be defined as (2) with* $w_i = m_p(i)$ *and* R *and* R^* *being defined as those in Proposition* 2.1. *Then* $(nb)^{-1}[(2/3)R(K_{r,m_p}) + (1/3)R^*(K_{r,m_p})]$, *the AIV of* $\hat{f}_r$ *based on weight function* m_p, *is non increasing in* p.

The proof of the above proposition is given in the Appendix. We now close this section with the following two remarks.

Remark 3.1 (Minimum variance weights). Given a weight function g that satisfies the assumptions (A1) and (A2), consider the mixture weight $m_p = (1 - p)g + p/r$, $p \in [0, 1]$. 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) those of $\hat{f}_r$ using g. Moreover, when $p = 1$, $\hat{f}_r$ has the smallest AIV when $g = g^U$, among the collection of all $\hat{f}_r$ based on the weight functions that satisfy the assumptions (A1) and (A2). Note that the assumption (A2) cannot be eliminated from Proposition 3.1. As a counterexample, for $r = 3$, let $g(1) = g(3) = 2/5$, $g(2) = 1/5$. In this case, $g^U(i) = 1/3$, $i = 1, 2, 3$. When $\hat{f}_r$ uses g, it follows that $\mathrm{AIV}(\hat{f}_r^U) = (8/27)/(nb) > \mathrm{AIV}(\hat{f}_r) = (22/75)/(nb)$. Hence, the discrete uniform weight function does not lead to minimum variance when assumption (A2) fails to hold.

Remark 3.2 (Minimum AMISE* of $\hat{f}_r^U$). The uniform weight $w_i = g^U(i), i = 1, 2, \ldots, r$, is the simplest weight and the minimum variance weight as discussed in Remark 3.1. Dong and Zheng (2001) showed that for even values of r, $\mathrm{AMISE}^*(\hat{f}_r^U)$ is minimized when $r = 6$. Figure 2 uses the upper solid curve to plot the $C(r)$ values of $\hat{f}_r^U$ for both odd and even values of r, which confirms Dong and Zheng's results, with $C(6) = 0.361334$. But the upper plus sign shows that the $\mathrm{AMISE}^*(\hat{f}_r^U)$ is minimized at $r = 5$ among all values of r, with $C(5) = 0.361262$.

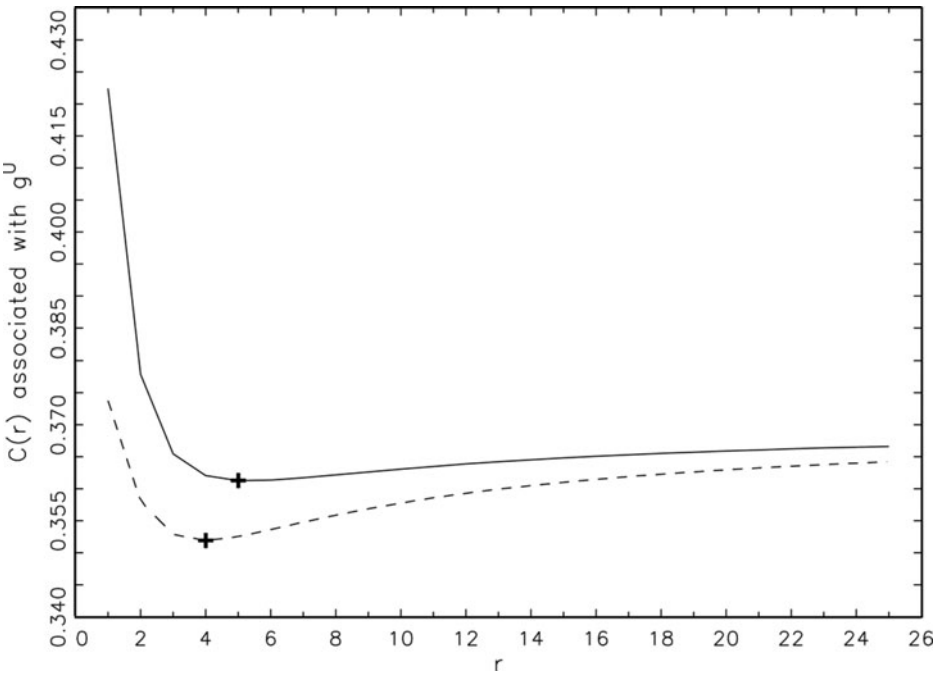

Figure 2. Values of $C(r)$ of $\hat{f}_r^U$ (solid curve) and $\tilde{f}_r^U$ (dashed curve). Plus signs denote their respective minimum values.

4. Frequency Polygon Based on Linearly Prebinned Data

In this section, we study the frequency polygon based on linearly prebinned data. The linearly prebinned data is defined by

$$\ell_k = \sum_{j=1}^{n} (1 - b^{-1}|X_j - t_k|)_+,$$

where the subscript "+" denotes the positive parts. This can be regarded as splitting the unit mass of each data point between the two nearest bin centers t_k and t_{k+1}, in reverse proportion to the distances to them. This rule has been proposed in Jones and Lotwick (1983) and is well known to be superior to the simple histogram binning rule, at least in the density estimation context. Let $\tilde{f}_r$ denote the frequency polygon based on the linear prebinned data $\ell_k' s$. To simplify the use of notations, we shall borrow the same notations used for $\hat{f}_r$ previously to express their counterparts used for $\tilde{f}_r$ if no confusion occurs in discussing the latter estimator.

4.1. The Weights for Linearly Prebinned Data

The weighted sum of the linearly prebinned data u_k and v_k used to construct $\tilde{f}_r$ by (1) is given by:

$$u_k = (1/nb) \sum_{j=1}^{r} w_j \ell_{k+j-(r+1)/2} = \frac{1}{n\, b(r+1)/2} \sum_{i=1}^{n} K_r \left(\frac{X_i - t_k}{b(r+1)/2} \right),$$

$$v_k = (1/nb)\sum_{j=1}^{r} w_j \ell_{k+j-(r+2)/2} = \frac{1}{n\,b(r+1)/2}\sum_{i=1}^{n} K_r\left(\frac{X_i - s_k}{b(r+1)/2}\right),$$

with the kernel function K_r is defined as

$$K_r(x) = \left[\frac{r+1}{2}\right]^2 \sum_{j=1}^{r+1}\left[w_{j-1}\left(\frac{2j}{r+1} - 1 - x\right) + w_j\left(-\frac{2(j-1)}{r+1} + 1 + x\right)\right] I_{E_j}, \quad (6)$$

where $w_0 \equiv w_{r+1} \equiv 0$ and $E_j = (-1+2(j-1)/(r+1), -1+2j/(r+1)]$. It can be shown that the current K_r is also a symmetric probability density supported on $[-1, 1]$. Therefore, u_k and v_k can be expressed as the kernel density estimates using bin width $b(r+1)/2$. The midpoint and edge frequency polygons, denoted by $\tilde{f}_r$ henceforth, are defined by (1) with u_k and v_k being replaced by the current ones, respectively. By such definition, $\tilde{f}_r$ is also a linear interpolant of kernel estimates. In the case of $r = 1$, the $\tilde{f}_1$ is the interpolated kernel density estimates (Jones, 1989) and also the kernel polygons of Lin et al. (2006) using triangular kernel function. As a direct result of Jones (1989), the following proposition is an analogue of Proposition 2.1.

Proposition 4.1. *Suppose that the density* f *and* g *satisfy the same conditions given in Proposition 2.1, respectively. If* $n \to \infty$, $b \to 0$ *and* $n\,b \to \infty$, *then the mean integrated square error formula of* $\tilde{f}_r$, $r \geq 1$, *is the same as (3), but with* $\mu_2 = \sum_{i=1}^{r} g(i)[(i - (r+1)/2)^2 + 1/6]$, $R(K_r) = (2/3)\sum_{i=1}^{r} g(i)^2 + (1/3)\sum_{i=1}^{r-1} g(i)g(i+1)$ *and* $R^*(K_r) = (2/3)\sum_{i=1}^{r-1} g(i)g(i+1) + (1/6)\sum_{i=1}^{r} g(i)^2 + (1/6)\sum_{i=1}^{r-2} g(i)g(i+2)$.

Note here that the AISB and AIV of $\tilde{f}_r$ are, respectively, given by $4^{-1}R(f'')b^4[\mu_2^2 + \mu_2/3 + 1/30]$ and $(n\,b)^{-1}[(2/3)R(K_r) + (1/3)R^*(K_r)]$. We have the following minimum AMISE of $\tilde{f}_r$ based on the optimal bin width

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

where, through a straightforward calculation,

$$C(r) = \left\{\mu_2^2 + \frac{\mu_2}{3} + \frac{1}{30}\right\}^{1/5}\left\{\frac{2}{3}R(K_r) + \frac{1}{3}R^*(K_r)\right\}^{4/5}.$$

We now end this subsection by the following remark.

Remark 4.1 (The performance of the proposed mixture weight function g_p). Let $C_p(r)$ denote the value of $C(r)$ when $\tilde{f}_r$ uses the weight function g_p and let $w_i = g_{p*}(i)$ denote the optimal mixture weight generated by (5) with the $p = p^*$ chosen from $[0, 1]$ to minimize the value of $C_p(r)$. Let w_i^* be the weights that minimize $C(r)$ among all choices of weights and let $\tilde{f}_r^*$ and $\tilde{f}_r^M$ denote $\tilde{f}_r$ using w_i^* and $w_i = g_{p*}(i)$, respectively. The values of w_i^* and $w_i = g_{p*}(i)$ (unreported but available upon request) obtained from exhaustive grid search show that the optimal w_i^* are very well approximated by our proposed $w_i = g_{p*}(i)$, for $r = 3, 4, \ldots, 7$. We compare the values of $C(r)$ and $C_p(r)$ of $\tilde{f}_r^*$ and $\tilde{f}_r^M$, respectively, by the lower plots of Fig. 1. It is seen in Fig. 1 that the $C_p(r)$ values of $\tilde{f}_r^M$ (lower dashed line) are almost the same as the $C(r)$ values of $\tilde{f}_r^*$ (lower plus signs) for $r = 3, 4, \ldots, 7$. Hence, the optimal weights and optimal performances of $\tilde{f}_r^*$ are very well approximated

by our proposed kernel method for the above values of r. Finally, as shown in Fig. 1, the linear binning rule leads to smaller AMISE values than simple binning rule and such gains diminish as the values of r increases.

4.2. The Minimum Variance Weight for Linearly Prebinned Data

In Sec. 3.2, we prove that the uniform weight leads to the minimum variance of $\hat{f}_r$. In this subsection we will use the following proposition to show that the use of uniform weight also minimizes the AIV of $\tilde{f}_r$. Its proof is given in the Appendix.

Proposition 4.2. *Given a non negative 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 (6) with* $w_i = m_p(i)$. *Then* $(nb)^{-1}[(2/3)R(K_{r,m_p}) + (1/3)R^*(K_{r,m_p})]$, *the AIV of* $\tilde{f}_r$ *using* m_p, *is non increasing in* p.

We now use the following two remarks to end this section.

Remark 4.2 (Minimum variance weight for $\tilde{f}_r$). By a similar argument as Remark 3.1, Proposition 4.2 implies that given a weight function g that satisfies (A1) and (A2), the AIV of $\tilde{f}_r$ based on m_p is smaller than (or equal to) that of $\tilde{f}_r$ using g. Moreover, Proposition 4.2 implies that $\tilde{f}_r$ has the smallest AIV when $g = g^U$, among the collection of all $\tilde{f}_r$ based on the weight functions that satisfy assumptions (A1) and (A2).

Remark 4.3 (Minimum AMISE* of $\tilde{f}_r^U$). The dashed curve in Fig. 2 plots the $C(r)$ values of the frequency polygon $\tilde{f}_r$ based on g^U (denoted by $\tilde{f}_r^U$). We conclude that AMISE*$(\tilde{f}_r^U)$ is minimized when $r = 4$, since $C(4) = 0.35190$ (lower plus sign) is the smallest among all values of $C(r)$.

5. Numerical Comparison

5.1. A Real Data Example

We illustrate $\hat{f}_r^*$, $r = 1, 2, \ldots, 7$, on a dataset consisting of 63 annual snowfalls (in inches) in Buffalo, New York, 1910–1972, given by Scott (1992) in which the data were applied to the author's proposed averaged shifted histogram. For a given bin width b, we follow Jones et al. (1998) to take $\min\{X_1, X_2, \ldots, X_{63}\} - b/10$ as our bin origin. Note here that the bin widths used in $\hat{f}_r^*$ for different values of r have a relation determined by formula (4). Therefore, the bin width, denoted by b_r, is given by $b_r = 11.3 \times h_r$, where the constants $h_r = b_r^*/b_1^* = 1$, 0.6986, 0.3661, 0.3216, 0.2893, 0.2644, 0.2445, respectively for $r = 1, 2, \ldots, 7$. We now use Fig. 3 to plot the resultant $\hat{f}_r^*$, $r = 1, 2, \ldots, 7$. For comparison's sake, each panel of Fig. 3 displays the density estimates $\hat{f}_r^*$ and $\hat{f}_{r+1}^*$, using the dashed and solid curve to plot the generalized midpoint and edge frequency polygons, respectively. For example, Fig. 3 a plots $\hat{f}_1^*$ and $\hat{f}_2^*$ by dashed and solid lines, respectively. It is seen from the figure that a bump to the right of the mode was masked by the larger bump at the mode of the EFP $\hat{f}_2^*$, while $\hat{f}_1^*$ and $\hat{f}_3^*$ is found to reveal such a bump. In general, the generalized midpoint frequency polygon reveals the three bumps better than the generalized edge frequency polygons when r is small. As the values of r increases the plots of $\hat{f}_r^*$ for both larger odd and even values of r reveal the presence of all three bumps. In addition, it tends to have smoother appearance while revealing three bumps as the value of r increases.

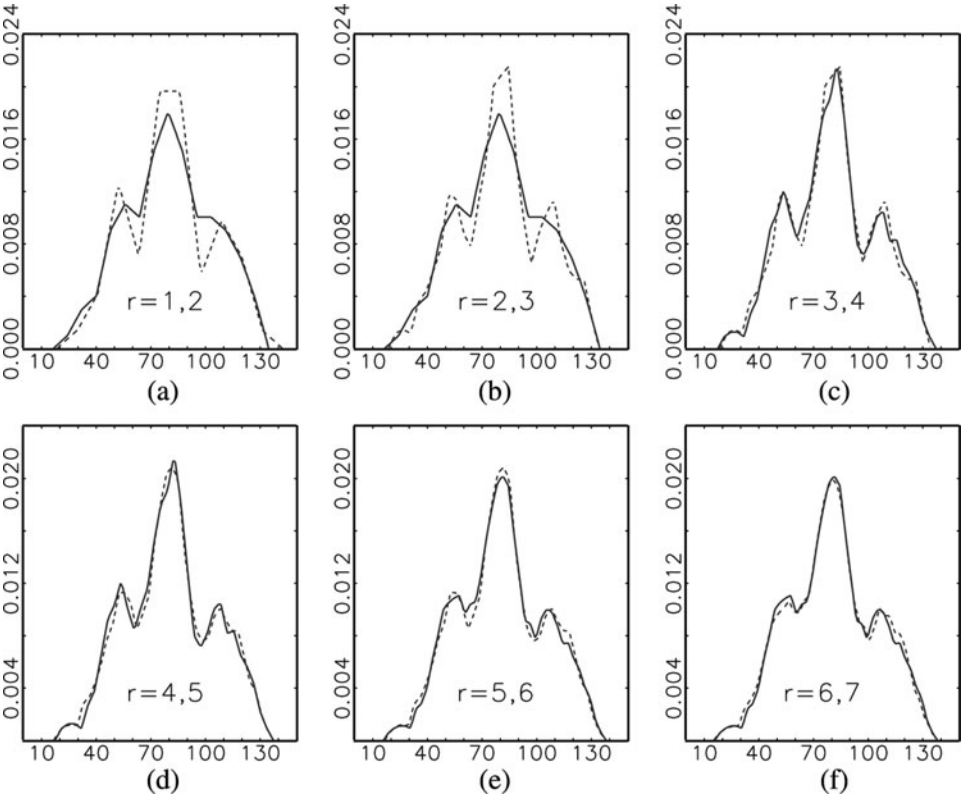

Figure 3. Plots of $\hat{f}_r^*$, $r = 1, 2, \ldots, 7$, applied to the snowfall data. In each panel the dashed and solid curves represent the generalized midpoint (r is odd) and edge frequency polygons (r is even), respectively.

As discussed in Sec. 3, $\hat{f}_r^*$ tend to be equally efficient when r exceeds 5. The estimated density plots in Figs. 3 d–f concur with this finding.

5.2. An Example with Simulated Data

In this subsection, we illustrate $\hat{f}_r^*$, $r = 1, 2, \ldots, 7$, on a simulated sample of size 100 generated from the standard normal distribution. For $r = 1, 2, \ldots, 7$, the bin width b_r used in $\hat{f}_r^*$ is given by $b_r = 1.08 \times h_r$, where the constants h_r are the same as those given in the previous subsection. We now use Fig. 4 to plot the resultant $\hat{f}_r^*$, $r = 1, 2, \ldots, 7$. In each panel of the figure, we use thick solid, dashed, and thin solid curves to plot the generalized midpoint frequency polygon, the generalized edge frequency polygon and the target standard normal density, respectively. It is seen in the figure that, as the value of r increases, $\hat{f}_r^*$ becomes smoother. This is because $\hat{f}_r^*$ join together more bin values to yield density estimates. Finally, Fig. 4 also concur with the previous finding that when the value of r exceeds 5, $\hat{f}_r^*$ tends to have similar accuracy.

In both the current and the previous examples, the generalized midpoint frequency polygon is found to be better at picking up the bumps of two densities. We now give the rationale as follows. When r is odd, the $([r/2] + 1)$th bin is the unique midmost bin and

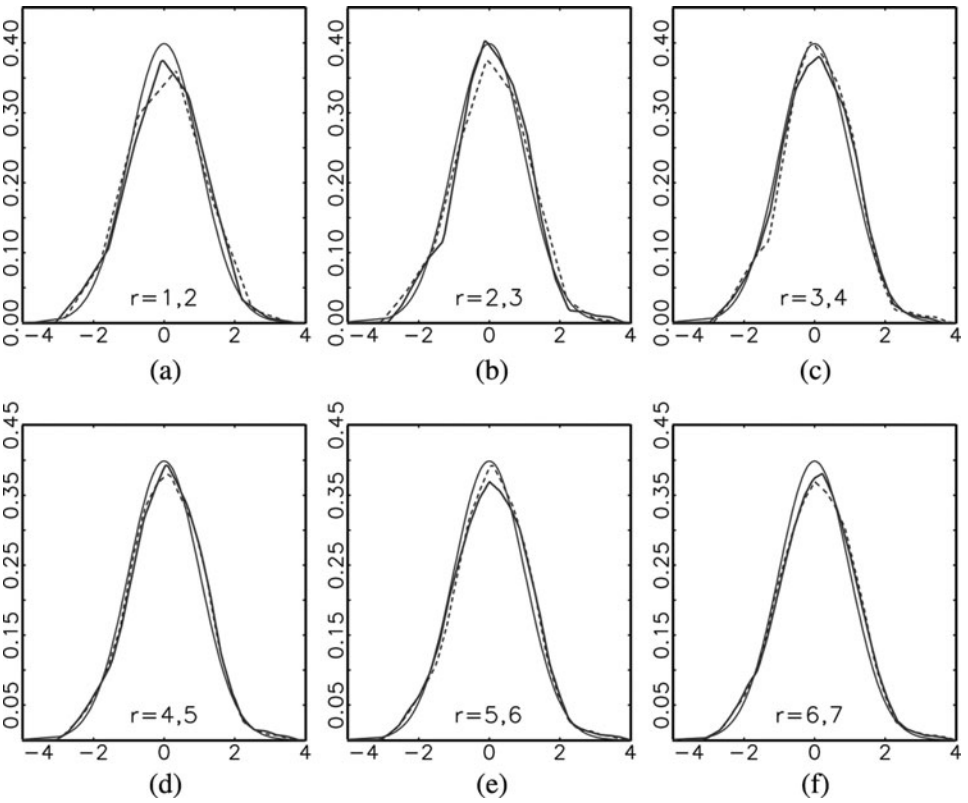

Figure 4. Plots of $\hat{f}_r^*, r = 1, 2, \ldots, 7$, applied to a simulated sample of size 100 drawn from standard normal distribution. In each panel the dashed and solid curves represent the generalized midpoint (r is odd) and edge frequency polygons (r is even), respectively.

the data lying therein is assigned the largest weight. When r is even, however, the $(r/2)$th and $(r/2 + 1)$th bins are the two midmost bins and they "share" the largest weights. As a consequence of such weighing schemes, the generalized midpoint frequency polygon tend to maintain the extremes while the generalized edge frequency polygon smoothes them out. This explains why midpoint frequency polygon is better at revealing the bumps of the densities, especially when the value of r is small.

5.3. Simulations

Here, 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; Dong and Zheng, 2001). The definitions of the eight densities are given at the end of this section. As pointed out in Jones et al. (1998), $\hat{f}_2^*$ enjoys 10% improvement over the ordinary midpoint frequency polygon $\hat{f}_1$ in terms of AMISE. Therefore, we will compare the performance of $\hat{f}_r^*, r = 2, 3, \ldots, 7$. To run the simulations, we generate random samples

$X_1, X_2, \ldots, X_n$ 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 $\min\{X_1, X_2, \ldots, X_n\} - b/10$ 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 1,000 repetitions of random samples from each density for each sample size. Then for each of $\hat{f}_r^*$, $r = 2, 3, \ldots, 7$, we numerically approximate each sample's respective integrated square error $\mathrm{ISE}(\hat{f}_r^*) = \int (\hat{f}_r^*(x) - f(x))^2 dx$ by $\hat{\mathrm{ISE}}(\hat{f}_r^*) = (1/400) \times \sum_{i=1}^{401} [\hat{f}_r^*(t_i) - f(t_i)]^2$, where $t_i = -4 + (i-1)(8/400)$, $i = 1, 2, \ldots, 401$. The simulated versions of these estimators' MISE are thus approximated by the average of the 1,000 sample's $\hat{ISE}$ values. Given the value of r, the MISE value of $\hat{f}_r^*$ is determined by the value of bin width b. For each set of 1,000 samples, we use 1,000 equally spaced grid values of bin width to computed their associated MISE values and obtain the minimal value of the simulated MISE. We denote the minimal MISE quantities of $\hat{f}_r^*$ by $\mathrm{MISE}^*(\hat{f}_r^*)$.

Table 3 reports the ratio $\mathrm{MISE}^*(\hat{f}_r^*)/\mathrm{MISE}^*(\hat{f}_2)$, $r = 3, 4, \ldots, 7$, obtained from the simulation results. The MISE ratios smaller than 1 indicate that the simulated minimal MISE's of $\hat{f}_r^*$ are smaller than that of EFP $\hat{f}_2$. Since the ratios for $\hat{f}_r^*$ in Table 3 is smaller than 1 for all sample sizes and all values of r, the frequency polygons $\hat{f}_r^*$, $r = 3, 4, \ldots, 7$, improve the accuracy of EFP for the above eight densities and the accuracy improves as r increases. The above results show that the improvements of $\hat{f}_r^*$, $r = 3, 4, \ldots, 7$, discussed in Sec. 3 in terms of their $C(r)$ quantities carry over to the situations of small and moderate sample sizes.

Remark 5.1 (The MISE ratios for $\hat{f}_r^M$). We have carried out an analogous Monte Carlo simulation to compare the performance of $\hat{f}_r^M$, $r = 2, 3, \ldots, 7$. The resultant MISE ratios $\mathrm{MISE}^*(\hat{f}_r^M)/\mathrm{MISE}^*(\hat{f}_2)$, $r = 3, 4, \ldots, 7$, are very close to those of $\hat{f}_r^*$ reported in Table 3, due to the closeness of optimal weights of $\hat{f}_r^*$ and the optimal mixture weights of $\hat{f}_r^M$. Hence the result are not reported here to save the space of this article.

Remark 5.2 (The MISE ratios for $\tilde{f}_r^*$). An analogous Monte Carlo simulation is also carried out to compare the performance of $\tilde{f}_r^*$, $r = 2, 3, \ldots, 7$. The resultant ratios $\mathrm{MISE}^*(\tilde{f}_r^*)/\mathrm{MISE}^*(\tilde{f}_2)$, $r = 3, 4, \ldots, 7$, are smaller than their respective counterpart $\mathrm{MISE}^*(\hat{f}_r^*)/\mathrm{MISE}^*(\hat{f}_2)$, $r = 3, 4, \ldots, 7$, and in general decreases as the value of r increases. This implies that the improvements of $\tilde{f}_r^*$, $r = 3, 4, \ldots, 7$, to $\tilde{f}_2$, in terms of their $C(r)$ quantities in Fig. 1 (lower plus sings) carry over to the situations of small and moderate sample sizes. As the simulation results parallel those reported in Table 3, they are not reported here but available upon request.

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: $0.1N(0, 1) + 0.9N(0, 0.1^2)$.
Bimodal: $0.5N(-1, (2/3)^2) + 0.5N(1, (2/3)^2)$.
Separated bimodal: $0.5N(-1.5, 0.5^2) + 0.5N(1.5, 0.5^2)$.
Skewed bimodal: $0.75N(0, 1) + 0.25N(1.5, (1/3)^2)$.
Trimodal: $0.45N(-1.2, 0.6^2) + 0.45N(1.2, 0.6^2) + 0.1N(0, 0.25^2)$.
Asymmetric claw: $0.5N(0, 1) + \sum_{\ell=-2}^{2} (2^{1-\ell}/31)N(\ell + 0.5, (2^{-\ell}/10)^2)$.

Table 3
Values of $\text{MISE}^*(\hat{f}_r^*)/\text{MISE}^*(\hat{f}_2)$, $r = 3, 4, 5, 6, 7$

	$n = 30$	$n = 50$	$n = 100$	$n = 200$
Gaussian				
3	0.973	0.956	0.948	0.956
4	0.948	0.941	0.936	0.937
5	0.940	0.924	0.918	0.926
6	0.937	0.923	0.918	0.920
7	0.925	0.916	0.913	0.917
Kurtotic				
3	0.982	0.977	0.970	0.964
4	0.974	0.966	0.958	0.954
5	0.968	0.959	0.952	0.947
6	0.965	0.956	0.949	0.944
7	0.962	0.954	0.946	0.942
Outlier				
3	0.965	0.955	0.962	0.962
4	0.944	0.935	0.948	0.945
5	0.933	0.929	0.937	0.937
6	0.931	0.920	0.930	0.931
7	0.928	0.921	0.930	0.928
Bimodal				
3	0.966	0.973	0.963	0.965
4	0.997	0.956	0.941	0.948
5	0.959	0.951	0.931	0.940
6	0.959	0.945	0.932	0.936
7	0.958	0.943	0.923	0.932
Separate bimodal				
3	0.952	0.962	0.972	0.961
4	0.949	0.950	0.953	0.946
5	0.939	0.940	0.940	0.938
6	0.934	0.932	0.937	0.932
7	0.932	0.932	0.934	0.930
Skewed Bimodal				
3	0.980	0.975	0.973	0.969
4	0.968	0.964	0.953	0.957
5	0.959	0.958	0.946	0.949
6	0.959	0.950	0.942	0.945
7	0.956	0.949	0.940	0.942
Trimodal				
3	0.981	0.971	0.972	0.976
4	0.971	0.966	0.959	0.963
5	0.968	0.962	0.954	0.955
6	0.968	0.961	0.947	0.952
7	0.965	0.956	0.945	0.949
Asymmetric claw				
3	0.987	0.991	0.989	0.989
4	0.984	0.982	0.983	0.979
5	0.974	0.976	0.977	0.975
6	0.971	0.974	0.974	0.972
7	0.968	0.972	0.972	0.971

Acknowledgments

The authors are very grateful to all referees, the Editor, and the Associate Editor for their insightful comments which help to improve the quality of this article.

Funding

The research of the first author was supported by the NSC under grant NSC-101-2118-M-032-001.

References

Dong, J. P., Zheng, C. (2001). Generalized edge frequency polygon for density estimation. *Statist. Probab. Lett*. 55:137–145.

Jones, M. C., Lotwick, H. W. (1983). On the errors involving in computing the empirical characteristic function. *J. Statist. Computat. Simul*. 17:133–149.

Jones, M. C. (1989). Discretized and interpolated kernel density estimates. *J. Amer. Statist. Assoc*. 84:733–741.

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

Lin, C. T., Wu, J. S., Yen, C. H. (2006). A note on kernel polygons. *Biometrika* 93:228–234.

Marron, J. S., Wand, M. P. (1992). Exact mean integrated squared error. *Ann. Statist*. 20:712–736.

Scott, D. W. (1985a). Frequency polygons: theory and application. *J. Amer. Statist. Assoc*. 80:348–354.

Scott, D. W. (1985b). Averaged shifted histograms: effective nonparametric density estimators in several dimensions. *Ann. Statist*. 13:1024–1040.

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

Scott, D. W. (2010). *Averaged Shifted Histograms*. Wiley Interdisciplinary Reviews: Computational Statistics. 2:160–164.

Simonoff, J. S., Udina, F. (1997). Measuring the stability of histogram appearance when the anchor position is changed. *Computat. Statist. Data Anal*. 23:335–353.

Appendix

Proof of Proposition 3.1. We now provide 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 therefore is omitted. There is nothing to prove for $r = 2$, since in this case $g(1) = g(2) = 1/2$ is the only choice of weight. Therefore the proof given below will be done for $r \geq 4$.

Given an arbitrary weight function g satisfying assumptions (A1) and (A2), let $V_g = 2R(K_r)/3 + R^*(K_r)/3$. It suffices to show that $(d/dp)V_{(1-p)g+p/r} \leq 0$. By Proposition 3.1 and the symmetry of g, one has $R(K_r) = \sum_{i=1}^{r}(g(i))^2$ and

$$\begin{aligned} R^*(K_r) &= \sum_{i=1}^{r-1} g(i)g(i+1) \\ &= \left[\sum_{i=1}^{r-1}(g(i))^2 + \sum_{i=1}^{r-1}(g(i+1))^2 - \sum_{i=1}^{r-1}(g(i+1)-g(i))^2\right]/2 \\ &= \sum_{i=1}^{r}(g(i))^2 - (g(r))^2 - \sum_{i=1}^{r/2-1}(g(i+1)-g(i))^2. \end{aligned}$$

Consequently,

$$V_g = 2R(K_r)/3 + R^*(K_r)/3 = \sum_{i=1}^{r}(g(i))^2 - \sum_{i=1}^{r/2-1}(g(i+1) - g(i))^2/3 - (g(r))^2/3.$$

By the definitions of m_p, $R(K_{r,m_p})$, $R^*(K_{r,m_p})$, and by replacing g with $m_p = (1-p)g + p/r$, the last formula implies

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

Differentiating both sides with respect to p yields

$$\begin{aligned}
(d/dp)V_{(1-p)g+p/r} &= -2(1-p)\left[\sum_{i=1}^{r}(g(i) - 1/r)^2 - \sum_{i=1}^{r/2-1}(g(i) - g(i+1))^2/3\right] \\
&\quad + 2((1-p)g(r) + p/r)(g(r) - 1/r)/3. \qquad (7)
\end{aligned}$$

The last term on the right-hand side of (7) is non positive, since $g(r) \le 1/r$ by assumptions (A1) and (A2). We now prove that the first term of (7) is also non-positive for $r \ge 4$. Let $j_r = \min_{1\le i\le r/2-1}\{i : g(i) \le 1/r \le g(i+1)\}$. It follows from (A2) that, for $1 \le i \le j_r - 1$, $(g(i) - 1/r)^2 \ge (g(i) - g(i+1))^2$, whereas for $j_r + 1 \le i \le r/2 - 1$, $(g(i+1) - 1/r)^2 \ge (g(i+1) - g(i))^2$. Consequently, the following inequality holds for the first term of (7)

$$\begin{aligned}
&-2(1-p)\left[\sum_{i=1}^{r}(g(i) - 1/r)^2 - \sum_{i=1}^{r/2-1}(g(i) - g(i+1))^2/3\right] \\
&\le -2(1-p)\left[2\sum_{i=1}^{r/2-1}(g(i) - 1/r)^2 - \sum_{i=1}^{r/2-1}(g(i) - g(i+1))^2/3\right] \\
&\le -2(1-p)\{\sum_{i=1}^{j_r-1}[(g(i) - 1/r)^2 - (g(i) - g(i+1))^2/3]
\end{aligned}$$

$$+\sum_{i=j_r+1}^{r/2-1}[(g(i+1)-1/r)^2-(g(i)-g(i+1))^2/3]$$
$$+(g(j_r)-1/r)^2+(g(j_r+1)-1/r)^2-(g(j_r)-g(j_r+1))^2/3\}$$
$$\le 0.$$

Hence, the proof of Proposition 3.1 is complete by noting that the sum of the last three terms in the above braces is positive, since $(b-x)^2+(a-x)^2-(a-b)^2/3 \ge 0$, whenever $0 < a \le x \le b$. □

Proof of Proposition 4.2. We now give the proof of Proposition 4.2 only for the case of even values of $r \ge 4$. The proof for the case of odd value of r is similar and thus is omitted. For a given weight function g satisfying assumptions (A1) and (A2), let $R(K_r)$ and $R^*(K_r)$ be those defined in Proposition 4.1 and let $V_g = 2R(K_r)/3 + R^*(K_r)/3$. Firstly, observe that

$$\sum_{i=1}^{r-1} g(i)g(i+1) = (1/2)\left[\sum_{i=1}^{r-1} g(i)^2 + \sum_{i=1}^{r-1} g(i+1)^2 + \sum_{i=1}^{r-1}(g(i+1)-g(i))^2\right]$$
$$= \sum_{i=1}^{r} g(i)^2 - \sum_{i=1}^{r/2-1}(g(i+1)-g(i))^2 - g(1)^2$$

and that, by the symmetry of g,

$$\sum_{i=1}^{r-2} g(i)g(i+2) = (1/2)\left[\sum_{i=1}^{r-2} g(i)^2 + \sum_{i=1}^{r-2} g(i+2)^2 + \sum_{i=1}^{r-2}(g(i+2)-g(i))^2\right]$$
$$= \sum_{i=1}^{r} g(i)^2 - \sum_{i=1}^{r/2-1}(g(i+2)-g(i))^2 - g(1)^2 - g(2)^2.$$

By the definition of $R(K_r)$, $R^*(K_r)$ and the above two formulae, V_g can be rewritten as

$$V_g = 2R(K_r)/3 + R^*(K_r)/3$$
$$= 4/9\sum_{i=1}^{r-1} g(i)g(i+1) + 1/2\sum_{i=1}^{r} g(i)^2 + 1/18\sum_{i=1}^{r-2} g(i)g(i+2)$$
$$= \sum_{i=1}^{r} g(i)^2 - (4/9)\sum_{i=1}^{r/2-1}(g(i+1)-g(i))^2 - (1/2)g(1)^2 - (1/18)g(2)^2$$
$$-(1/18)\sum_{i=1}^{r/2-1}(g(i+2)-g(i))^2. \tag{8}$$

Then (8) and the following formula

$$\sum_{i=1}^{r}((1-p)g(i)+p/r)^2 = (1-p)^2\sum_{i=1}^{r}(g(i)-1/r)^2 + 1/r$$

imply that for the weight function $m_p = (1-p)g + p/r$, $p \in [0,1]$, $V_{m_p} = V_{(1-p)g+p/r}$ can be expressed by

$$\begin{aligned} V_{(1-p)g+p/r} &= 2R(K_{r,m_p})/3 + R^*(K_{r,m_p})/3 \\ &= (1-p)^2 \sum_{i=1}^{r} (g(i)-1/r)^2 + 1/r - (4/9)(1-p)^2 \sum_{i=1}^{r/2-1} (g(i+1)-g(i))^2 \\ &\quad - (1/18)(1-p)^2 \sum_{i=1}^{r/2-1} (g(i+2)-g(i))^2 \\ &\quad - (1/2)(1-p)^2(g(1)-1/r)^2 - (1/18)(1-p)^2(g(2)-1/r)^2 \\ &\quad - (1-p)(g(1)-1/r)/r - (1/9)(1-p)(g(2)-1/r)/r - (5/9)r^2. \end{aligned}$$

Differentiating both sides of the above equation with respect to p yields

$$\begin{aligned} (d/dp)V_{(1-p)g+p/r} = \Bigg[& -2(1-p)\sum_{i=1}^{r}(g(i)-1/r)^2 \\ & + (8/9)(1-p)\sum_{i=1}^{r/2-1}(g(i+1)-g(i))^2 \\ & + (1/9)(1-p)\sum_{i=1}^{r/2-1}(g(i+2)-g(i))^2 \\ & + (1-p)(g(1)-1/r)^2 + (1/9)(1-p)(g(2)-1/r)^2 \Bigg] \\ & + (g(1)-1/r)/r + (1/9)(g(2)-1/r)/r. \end{aligned} \tag{9}$$

To prove Proposition 4.2, it suffices to show that $(d/dp)V_{(1-p)g+p/r} \le 0$, for $r \ge 4$. Firstly, note that the sum of the last two terms of (9) is non positive. To see this,

$$\begin{aligned} 1 &= \sum_{i=1}^{r} g(i) = 2g(1) + \sum_{i=2}^{r-1} g(i) \ge 2g(1) + (r-2)g(2) \\ &\Rightarrow g(2) - \frac{1}{r} \le \frac{1-2g(1)}{r-2} - \frac{1}{r} = \frac{2}{r-2}\Big(\frac{1}{r} - g(1)\Big), \end{aligned}$$

Consequently, it holds that

$$[g(1)-1/r] + \frac{1}{9}[g(2)-1/r] \le [g(1)-1/r][1-(2/9)(r-2)^{-1}] \le 0. \tag{10}$$

Next, to complete the proof of Proposition 4.2, we now prove that the sum of the remaining terms in the brackets of (9) is also non-positive. Let $j_r = \min_{1\le i\le r/2-1}\{i : g(i) \le 1/r \le g(i+1)\}$. As asserted earlier, for $1 \le i \le j_r - 1$, $(g(i)-1/r)^2 \ge (g(i)-g(i+1))^2$, whereas for $j_r+1 \le i \le r/2-1$, $(g(i+1)-1/r)^2 \ge (g(i+1)-g(i))^2$. The following three formulae

(11)–(13) are needed in completing the proof of Proposition 4.2:

$$\text{(i)}\quad \sum_{i=1}^{r}(g(i)-1/r)^2 = 2\sum_{i=1}^{r/2}(g(i)-1/r)^2, \tag{11}$$

$$\begin{aligned}\text{(ii)}\quad &\sum_{i=1}^{r/2-1}[g(i+1)-g(i)]^2\\ &\le \sum_{i=1}^{r/2}[g(i)-1/r]^2 + 2[g(j_r+1)-1/r][1/r-g(j_r)]\\ &\le \sum_{i=1}^{r/2}[g(i)-1/r]^2 + [g(j_r+1)-1/r]^2 + [g(j_r)-1/r]^2,\end{aligned} \tag{12}$$

and

$$\begin{aligned}\text{(iii)}\quad &\sum_{i=1}^{r/2-1}[g(i+2)-g(i)]^2\\ &\le \sum_{i=1}^{r/2+1}[g(i)-1/r]^2 + 2[g(j_r+2)-1/r][1/r-g(j_r)]\\ &\quad + 2[g(j_r+1)-1/r][1/r-g(j_r-1)]I_{\{j_r\neq 1\}}\\ &\le \sum_{i=1}^{r/2+1}[g(i)-1/r]^2 + [g(j_r+2)-1/r]^2 + [g(j_r)-1/r]^2\\ &\quad + \{[g(j_r+1)-1/r]^2 + [g(j_r-1)-1/r]^2\}I_{\{j_r\neq 1\}}.\end{aligned} \tag{13}$$

Combining (12), (13), and that $g(r/2) = g(r/2-1)$ due to symmetry, we have

$$\begin{aligned}&(8/9)(1-p)\sum_{i=1}^{r/2-1}[g(i+1)-g(i)]^2 + (1/9)(1-p)\sum_{i=1}^{r/2-1}[g(i+2)-g(i)]^2\\ &\le (1-p)\left\{\sum_{i=1}^{r/2}(g(i)-1/r)^2 + (1/9)[g(r/2)-1/r]^2\right.\\ &\quad + [g(j_r)-1/r]^2 + [8/9+(1/9)I_{\{j_r\neq 1\}}][g(j_r+1)-1/r]^2\\ &\quad \left.+ (1/9)[g(j_r+2)-1/r]^2 + (1/9)[g(j_r-1)-1/r]^2 I_{\{j_r\neq 1\}}\right\}.\end{aligned} \tag{14}$$

By using (9), (10), (11), and (14), we have

$$\begin{aligned}(d/dp)V_{(1-p)g+p/r} \le -(1-p)\{&3\left[\sum_{i=1}^{r/2-1}(g(i)-1/r)^2\right] + (26/9)(g(r/2)-1/r)^2\\ &- [g(j_r)-1/r]^2 - [8/9+(1/9)I_{\{j_r\neq 1\}}][g(j_r+1)-1/r]^2\end{aligned}$$

$$
\begin{aligned}
&- (1/9)[g(j_r + 2) - 1/r]^2 - (1/9)[g(j_r - 1) - 1/r]^2 I_{\{j_r \neq 1\}} \\
&- (g(1) - 1/r)^2 - (1/9)(g(2) - 1/r)^2 \Big\} \\
&+ [g(1) - 1/r][1 - (2/9)(r-2)^{-1}] \\
&\leq 0,
\end{aligned}
$$

Hence, the proof of Proposition 4.2 is complete by noting that in the above braces the first term contains all terms that appear later with minus sign. □