Smoothed Histograms for Frequency Data on Irregular Intervals

David W. Scott and Warren R. Scott

Frequency tables are often constructed on intervals of irregular width. When plotted as bar charts, the underlying true density information may be quite distorted. The majority of introductory statistics texts recommend tabulating data into intervals of equal width, but seldom caution the consequences of failing to do so. An occasional introductory text correctly emphasizes that area rather than frequency should be plotted. Nevertheless, the correctly scaled density figure is often visually less informative than one might expect, with wide bins at constant height. In many cases, the rightmost bin interval has no well-defined endpoint, making its depiction somewhat arbitrary. In this note, we introduce a regular histogram approximation that matches the frequencies and also minimizes a roughness criterion for visual and exploratory appeal. The resulting estimate can reveal the density structure much more clearly. We also formulate an alternative criterion that explicitly takes account of the uncertainty in the bin frequencies.

KEY WORDS: Density approximation; Unequal bin intervals; Penalized roughness criterion.

1. INTRODUCTION

The data snapshot feature of *USA TODAY* on Friday, October 13, 2006 displayed frequency data taken from a survey on marriage; see Figure 1. This bar chart exemplifies the difficulties with irregular intervals or bins. The first two bins are six months wide, the third, two years wide, and the final bin width is not specified. In addition, the percentages in the figure cannot all be correct as the sum is only 90%. Where is the missing 10%?

A check of the original source online shows a table (not a bar chart) with the percentages as given. But the table also provides the raw bin counts of 181, 147, 651, and 228 for a total sample of 1,207. Thus the first bin has a frequency of 15.0% not 5.0% as shown, a typo which accounts for the missing 10%.

David W. Scott is Noah Harding Professor, Department of Statistics MS-138, Rice University, Houston, TX 77251-1892 (E-mail: *scottdw@rice.edu*). Warren R. Scott is a Graduate Student, Department of Operations Research and Financial Engineering, Princeton University, Engineering Quadrangle, Princeton, NJ 08544 (E-mail: *wscott@princeton.edu*). This work was partially supported by the sabbatical program of the National Security Agency, NSF award DMS-05-05584, and ONR contract N00014-06-1-0060.

In order to preserve areas, each bin frequency must be divided by its bin width (in years); see Section 2. In Figure 2, we display two proper density histograms, one with the final bin chosen to be of width two years, and a second of width four years. These are visually quite different from each other, but the true width of the final bin is unknown. Comparing Figures 1 and 2, the third bin height is much lower in Figure 2 and nearly identical to the second bin height when properly normalized. In other words, since the third bin is four times wider than the second bin, the third bin height is plotted four times higher than it should be in Figure 1.

Can a more informative histogram be constructed with such limited data? If we make the assumption that the true underlying density is smooth and continuous, then the answer is often yes. We investigate this in the following.

2. A PROBABILITY HISTOGRAM FOR IRREGULAR BINS

Let us introduce notation for a true density histogram with M irregular intervals with endpoints $t_0 < t_1 < \cdots < t_M$; see

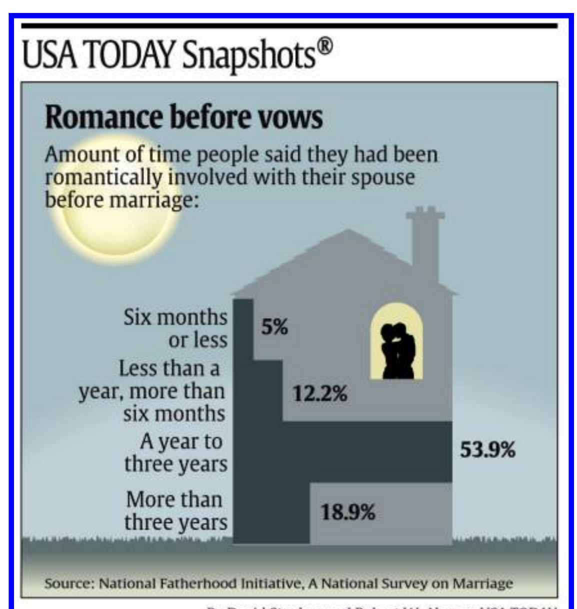

Figure 1. Original USA TODAY "Snapshot" front page art. The four-bin histogram is shown as a horizontal bar chart. Note the four frequencies total only 90%. ©*USA TODAY*. October 13, 2006. Reprinted with Permission.

 DOI: 10.1198/000313008X335581

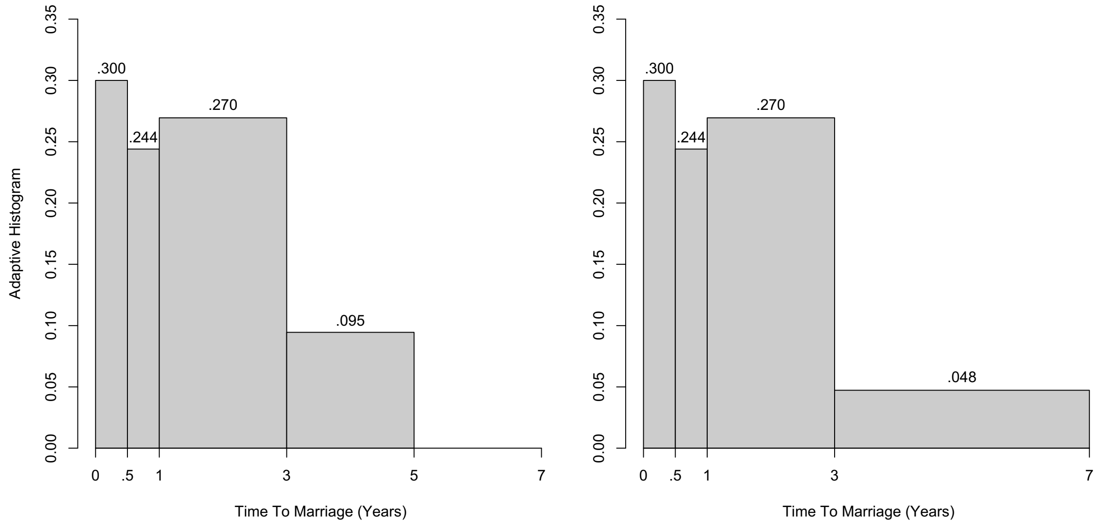

Figure 2. Two proper adaptive density histograms of the corrected marriage data, with the first interval representing 15.0% rather than 5% of the total. The last open-ended bin is shown with widths 2 and 4 years. However, one can cite isolated examples where this interval might be over 20 years wide, in which case the height would be very close to zero over a long interval.

Figure 3. Let the kth bin be denoted by $B_k = [t_{k-1}, t_k)$ with bin width denoted by $h_k = t_k - t_{k-1}$, $k = 1, 2, \ldots, M$. Let n_k and f_k denote the sample size and relative frequency, respectively, in bin B_k. Then the total sample size is $n = \sum_{k=1}^{M} n_k$ and $f_k = n_k/n$, $k = 1, 2, \ldots, M$, so that $\sum f_k = 1$. Plotting a bar chart with heights proportional to n_k/h_k or f_k/h_k gives the proper relative areas; however, the latter choice defines a density histogram:

$$\hat{f}(x) \equiv \frac{f_k}{h_k} = \frac{n_k}{nh_k}, \quad x \in B_k$$

as it is easy to check that $\int \hat{f}(x)dx = 1$ and $\hat{f}(x) \geq 0$.

3. AREA-PRESERVING SMOOTHED HISTOGRAM

We propose to construct an *equally spaced* probability histogram, denoted by $g(x)$, with smaller bin width δ to provide a superior density estimate, at least for visualization and exploration purposes. Each of the h_k's should be an integer multiple of δ. Denote the integer multiple by $m_k = h_k/\delta$. For example, with the *USA TODAY* data, δ could be $1/2$ or $1/8$ year.

Note that $g(x)$ is defined on the interval $[t_0, t_M)$. Clearly the total number of bins (of width δ) in $g(x)$ equals

$$m \equiv \sum_{k=1}^{M} m_k = \frac{t_M - t_0}{\delta}.$$

Denote the values of $g(x)$ in the new narrower bins of width δ by

$$g(x) \equiv g_\ell, \quad x \in I_\ell \equiv [t_0 + (\ell - 1)\delta, t_0 + \ell\delta), \quad \ell = 1, 2, \ldots, m.$$

We define $g(x)$ to be 0 outside the interval $[t_0, t_M)$, and $g_\ell = 0$ there, too.

We require that the new histogram, $g(x)$, match the original bin frequencies, that is,

$$\int_{B_k} g(x)\,\mathrm{d}x = f_k, \quad k = 1, 2, \ldots, M. \tag{1}$$

These M constraints do not uniquely determine $g(x)$. Since we assume the true density is continuous and smooth, we seek a $g(x)$ that is similarly visually pleasing. From cubic spline theory, we know that curves which minimize the roughness (R)

$$R(g) \equiv \int_{-\infty}^{\infty} g''(x)^2\,\mathrm{d}x \tag{2}$$

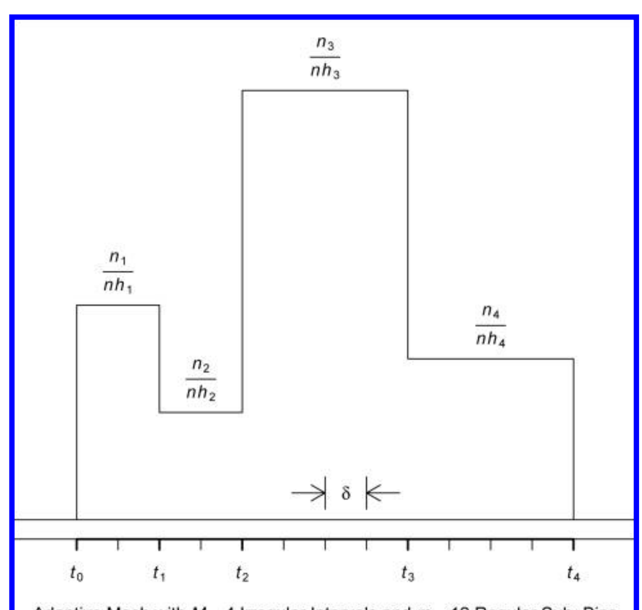

Figure 3. Notation for proper adaptive histogram on irregular frequency intervals. The new proposed smoothed histogram uses a finer mesh that is equally spaced with bin width δ.

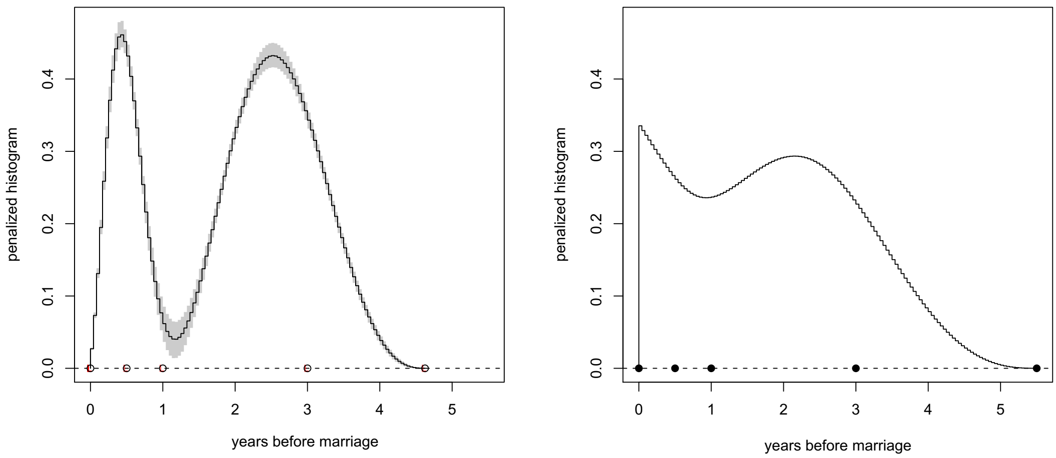

Figure 4. (left frame) The proposed frequency-matching smoothed histogram density estimate, with $\delta = 1/24$ year (half a month). The width of the final interval (1.625 years) was determined as the largest value where the density estimate was nonnegative. The 2 modes are at 0.43 and 2.52 years. The points on the x-axis indicate the original bin boundaries. The shaded area was obtained from bootstrap samples; see text. (right frame) The smoothed histogram that is allowed to "float" on the left side; see text.

are smooth and visually attractive; see de Boor (1979). We wish to use this function in our criterion.

Now our new histogram $g(x)$ is piecewise constant so that $g''(x)$ is not well-defined. However, by using finite differences and choosing δ small, we can develop a useful approximation for the smoothing penalty function in Equation (2). From ordinary calculus, $(g_{\ell+1} - g_\ell)/\delta$ gives an approximation to the first derivative of $g(x)$ and $(g_{\ell+1} - 2g_\ell + g_{\ell-1})/\delta^2$ gives an approximation to the second derivative in bin I_ℓ. Thus our discrete approximation to the roughness function in Equation (2) is

$$R(g) \approx \sum_{\ell=-\infty}^{\infty} \left(\frac{g_{\ell+1} - 2g_\ell + g_{\ell-1}}{\delta^2} \right)^2 \cdot \delta , \qquad (3)$$

which includes the 0 levels outside $[t_0, t_M)$. Define the vector

$$\mathbf{g} = (g_1\ g_2\ \dots\ g_m)^T ;$$

then it is easy to verify that $R(g) = \mathbf{g}^T \mathbf{A} \mathbf{g}$, which is a quadratic form, where $\mathbf{A} = (a_{ij})$ is an $m \times m$ symmetric matrix of 0's except with $a_{ii} = 6$, $a_{i(i+1)} = a_{(i+1)i} = -4$, and $a_{i(i+2)} = a_{(i+2)i} = 1$, for indices in the range from 1 to m, $m-1$, and $m-2$, respectively.

The frequency constraints in Equation (1) are simple to implement. For example in bin B_1,

$$\sum_{\ell=1}^{m_1} g_\ell \cdot \delta = f_1 ,$$

and for the second bin

$$\sum_{\ell=m_1+1}^{m_1+m_2} g_\ell \cdot \delta = f_2 \quad \text{etc.}$$

Let $\mathbf{0}_n$ and $\mathbf{1}_n$ denote *column* vectors of 0's and 1's of length n. In matrix form, the M bin frequency constraints may be written as

$$\begin{bmatrix} \delta \cdot \mathbf{1}_{m_1}^T & \mathbf{0}_{m_2}^T & \mathbf{0}_{m_3}^T & \dots & \mathbf{0}_{m_M}^T \\ \mathbf{0}_{m_1}^T & \delta \cdot \mathbf{1}_{m_2}^T & \mathbf{0}_{m_3}^T & \dots & \mathbf{0}_{m_M}^T \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{0}_{m_1}^T & \mathbf{0}_{m_2}^T & \mathbf{0}_{m_3}^T & \dots & \delta \cdot \mathbf{1}_{m_M}^T \end{bmatrix} \cdot \mathbf{g} = \begin{bmatrix} f_1 \\ f_2 \\ \vdots \\ f_M \end{bmatrix} , \qquad (4)$$

say $\mathbf{Wg} = \mathbf{f}$, where $\mathbf{W}$ is the $M \times m$ matrix on the left of Equation (4) and $\mathbf{f} = (f_1\ f_2\ \dots\ f_M)^T$.

Thus, our proposed smoothed histogram solves the following constrained optimization problem:

$$\hat{\mathbf{g}} = \arg\min_{\mathbf{g}}\ \mathbf{g}^T \mathbf{A} \mathbf{g} \qquad s/t \quad \mathbf{Wg} - \mathbf{f} = \mathbf{0_M} . \qquad (5)$$

We do not explicitly impose $\mathbf{g} \geq \mathbf{0_m}$ here. In the Appendix, we show that

$$\hat{\mathbf{g}} = \mathbf{A}^{-1} \mathbf{W}^T (\mathbf{W} \mathbf{A}^{-1} \mathbf{W}^T)^{-1} \mathbf{f}$$

(assuming $\mathbf{A}^{-1}$ exists, which it does). Note that $\mathbf{W}\hat{\mathbf{g}} = \mathbf{f}$ as desired.

In the left frame of Figure 4, the smoothed penalized histogram for the marriage data with $\delta = 1/24$ year (half a month) is shown. The mildly bimodal shape suggested in Figure 2 is much more strongly indicated in Figure 4, with widely separated modes at 0.43 and 2.52 years. In order to minimize $R(g)$ with $g \geq 0$, the right endpoint $t_M = 4.67$ years ($m = 111$ bins) was selected to be as large as possible so that all $g_\ell \geq 0$; otherwise, some g_ℓ are negative. The nonnegativity constraint can also be handled more elegantly using the quadratic programming routine, `quadprog`, which is available in R. Our software is available in R at the first author's Web site, *www.stat.rice.edu/~scottdw/TAS/*, together with the code that generated the figures in this article.

Now it is likely that there are a handful of individual cases beyond the right boundary, t_M. But we take our estimate as only an approximation and not overly precise in the absence of further frequency information.

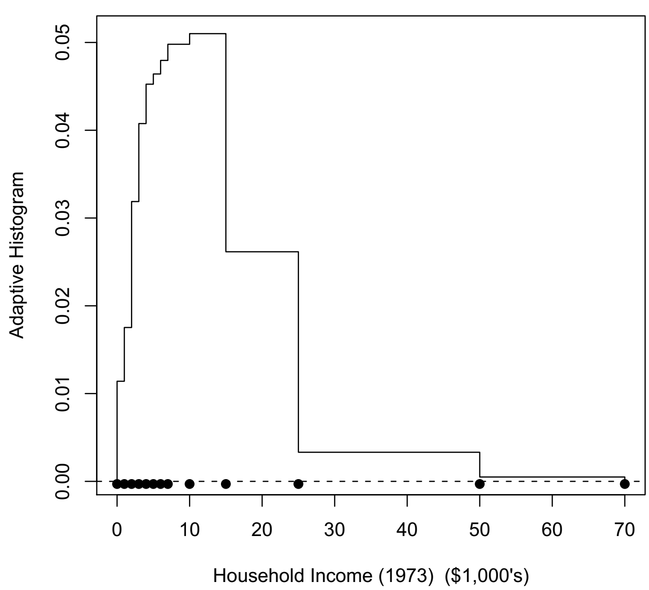

Figure 5. Proper adaptive histogram of the 1973 Family Household Income survey. The bin widths follow the choice by Freedman et. al. The final bin is shown with arbitrary width of \$20,000.

To examine the stability and reproducibility of our estimate, we drew 1,000 bootstrap samples with $n = 1207$ from a multinomial using the four sample bin frequencies. Each of the resulting 1000 area-matching estimates was also strongly bimodal. The shaded area in Figure 4 shows the 25th and 75th percentiles of these bootstrap estimates.

4. ALTERNATIVE BOUNDARY CHOICES

The choice of penalty in Equation (3) assumes that the histogram $g(x)$ should smoothly approach 0 on both ends. Sometimes you may wish to let $g(x)$ "float" on the left side. This is achieved by starting the penalty at t_0 rather than including the zero values for $t < t_0$. The change to $\mathbf{A}$ is quite simple: set the four values $a_{11} = 1$, $a_{22} = 5$, $a_{12} = -2$, and $a_{21} = -2$ ($\mathbf{A}$ is still invertible). If the right end is to "float", then set $a_{m,m} = 1$, $a_{m-1,m-1} = 5$, $a_{m,m-1} = -2$, and $a_{m-1,m} = -2$. Or if you want both ends to float, make both sets of changes. (However, $\mathbf{A}$ is no longer invertible. A more general solution is given in the Appendix.)

The right frame of Figure 4 shows the estimate with the left end "floating." This estimate spikes at the origin. But intuitively, not that many individuals elope on the first day they meet, and thus we prefer the left frame in Figure 4.

5. THE 1973 HOUSEHOLD INCOME EXAMPLE

One introductory textbook that emphasizes the importance of area versus frequency is *Statistics* by Freedman et al. (1978). As an example, they consider the 1973 household income from the Department of Commerce publication, Current Population Report Series P-60 #97 (available online). The bin boundaries (in \$1000's) are $\{0, 1, 2, 3, 4, 5, 6, 7, 10, 15, 25, 50, > 50\}$ and frequencies (in percentages)

$$f_k = (1.1,\ 1.8,\ 3.2,\ 4.1,\ 4.5,\ 4.6,\ 4.8,\ 14.9,\ 25.5,\ 26.2,\ 8.3,\ 1.0)^T$$

based upon a sample of approximately 40,000 according to the authors; see Figure 5. Choosing $\delta = \$500$, Figure 6 displays our smoothed histogram estimate (with the right most end point chosen to keep $\hat{g}_\ell \geq 0$). Again a bimodal feature is suggested. Observe that the estimates seems a little rough for $x < \$15K$ due to more noise in the narrow bins.

Exactly matching bin frequencies does not make good statistical sense if their accuracies do not warrant it. In the next section, we propose an alternative criterion that works well in this setting.

6. SMOOTHED HISTOGRAM FOR NOISY FREQUENCIES

If M is large and/or n is small, then some of the frequencies f_k are likely to be much more accurate than others. If the sample size is known, then $\text{var}(f_k) = f_k(1-f_k)/n$. Let $\mathbf{V}$ be the $M \times M$ diagonal matrix with these variances.

Rather than insisting $\mathbf{Wg} = \mathbf{f}$ exactly, we allow a discrepancy of $(\mathbf{Wg}-\mathbf{f})^T\mathbf{V}^{-1}(\mathbf{Wg}-\mathbf{f})$ so that the $\hat{g}_\ell$ will more closely match frequencies that are more accurate than those that are not. We still wish to enforce smoothness by augmenting with the same roughness penalty. Our new criterion becomes

$$\hat{\mathbf{g}} = \arg\min_{\mathbf{g}} (\mathbf{Wg}-\mathbf{f})^T\mathbf{V}^{-1}(\mathbf{Wg}-\mathbf{f}) + \lambda \mathbf{g}^T\mathbf{Ag}, \qquad (6)$$

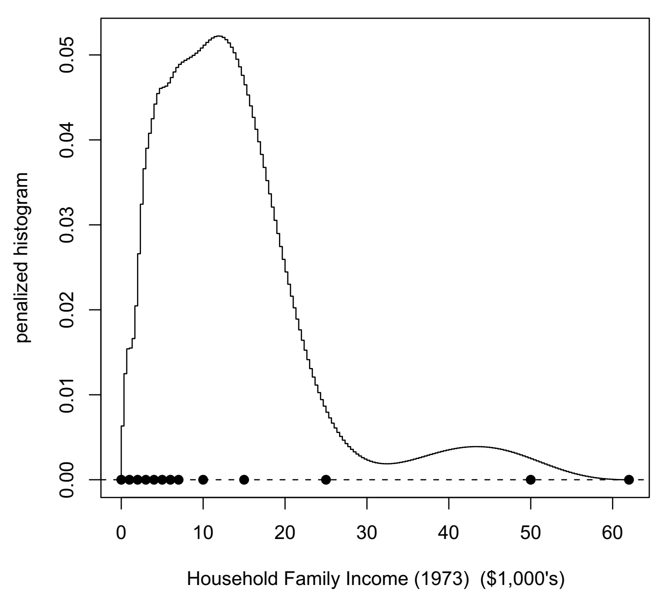

Figure 6. The frequency-matching penalized histogram shows an interesting bimodal structure. However, the estimate for $x < \$15,000$ reveals that the noise in the bin frequencies should not be ignored. The right endpoint was again chosen as the largest value where the estimate is nonnegative.

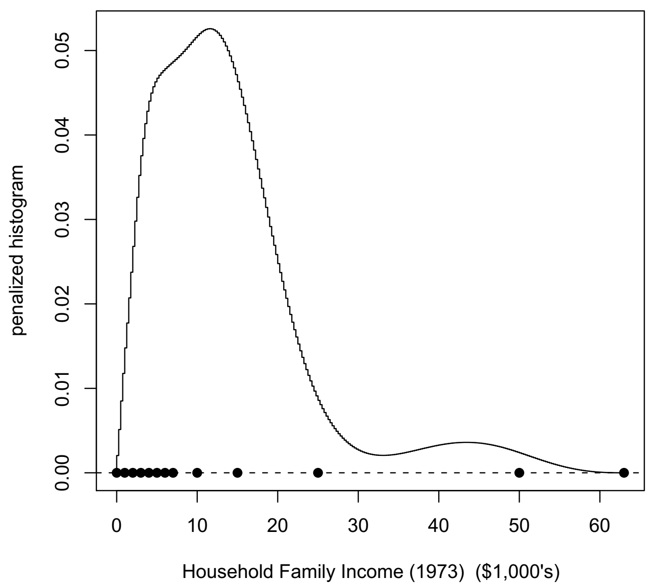

Figure 7. The second formulation of the penalized histogram that explicitly takes into account the noise in the interval frequencies. The basic structure is identical to that in Figure 6, but is visually superior.

where λ is a regularization or penalty parameter. We have only one overall constraint, namely that $g(x)$ integrate to one, which may be expressed as $\mathbf{g}^T \mathbf{1}_m = 1/\delta$.

The solution to this problem is also given in the Appendix. An example for the income data is shown in Figure 7. Notice how the visual roughness for $x < \$15K$ has been removed.

We initially chose values for the penalty λ interactively by eye. Rudemo (1982) described a cross-validation algorithm for histograms,

$$\hat{\lambda} = \arg\min_{\lambda} \int \hat{g}_\lambda(x)^2\, dx - \frac{2}{n} \sum_{i=1}^{n} \hat{g}_{\lambda,-i}(x_i)\,, \qquad (7)$$

where $\hat{g}_{\lambda,-i}(x_i)$ is the smoothed histogram estimate with penalty parameter λ based on a sample of size $n-1$ with x_i deleted, then evaluated at $x = x_i$. Note that since $\hat{g}_{\lambda,-i}(x)$ is constant over each bin of width δ, $\hat{g}_{\lambda,-i}(x_i)$ is identical for all x_i in the same bin. Therefore, the second term in Equation (7) simplifies greatly. However, Equation (7) requires counts for the bin width δ, which we do not have. Instead, we tried using an average of the penalized histogram over the original bin widths h_k, and this gave very similar choices to our subjective ones. The value used in Figure 7 was $\lambda = 10^{5.3}$.

7. DISCUSSION AND RELATED WORK

In this note, we have found a useful histogram approximation when the raw data are frequencies from bins of unequal width, and when the final bin width is unknown. While a density histogram is technically correct and superior to the incorrect bar chart such as that displayed in the *USA TODAY* feature, our smoothed area-matching histogram can provide a superior estimator for exploratory and visual understanding of the true underlying density. We were able to gain resolution by making two assumptions: the true density should be smooth, and the new smaller bins should exactly combine to the larger original bins. Since matching bin frequencies does not uniquely specify an estimator, we found the histogram-like estimator that minimizes a discrete version of the penalty function commonly used to derive cubic splines. We also introduced a variation on the area-preserving estimator that accounts for noise in the bin frequencies. This version does not exactly match bin frequencies, but trades off the discrepancy against the same penalty function, while constraining the estimator to integrate to one. An approximation to a cross-validation function is introduced to help guide the choice of the weight on the penalty function.

These smoothed histograms were applied to two datasets. With the marriage and income datasets, we have been able to turn coarse histograms with ill-defined right endpoints into more informative histograms that are bimodal. We believe these bimodal features are real, but as Wainer (2006) has noted, some irregular meshes may result in artifacts, or may miss smaller features.

We briefly review other work that led us to propose the two estimators investigated in this note. The first explicit results on optimal histograms with equally spaced bin widths were given by Scott (1979) and Freedman and Diaconis (1981). Extensive work on data-based versions were surveyed by Scott and Terrell (1987) and Wand (1997). Determining an optimal data-based irregular mesh is much more difficult in practice than finding an optimal fixed-width mesh. Simple transformation seems a good strategy to employ at this time. For example, a version of Figure 7 with the data transformed to a log-income scale is also bimodal, but the right-hand endpoint corresponds to \$144,000 (not shown). This density is nearly symmetrical.

There is an extensive literature on area matching and penalized density estimation, which we briefly review. Most approaches require complete rather than binned data. Boneva et al. (1971) described a continuous spline with first-derivative penalty, which matches the bin areas of an equally spaced histogram. Matching areas between selected sample percentiles with cubic splines was described by Scott (1979, pp. 177–179). Minnotte (1996) defined a frequency polygon that matches regular histogram frequencies in each bin. Scott and Sagae (1997) matched not only frequencies but also other bin moments with low-order polynomials. By contrast, our estimate is explicitly in the form of a histogram.

Penalized density estimation also has an interesting history. Good and Gaskins (1980) used penalized likelihood to find bumps. Scott et al. (1980) used a discrete penalty and maximum likelihood to find a histogram-like estimator. But most work has focused on the use of spline approximations with various penalties on maximum likelihood: Wahba (1981) penalized the density itself; Kooperberg and Stone (1991) focused on the log-density; and Eilers and Marx (1996) introduced the P-spline by defining their penalty not on the density but on its spline coefficients. These are representative of likelihood-based methods. One interesting variation is the linear estimator of Terrell (1990) that adds a penalty to the criterion of Rudemo (1982) given in Equation (7).

Our second estimator is close in spirit to the density estimator given by Eilers and Marx (1996), who viewed density estimation as a regression smoother of a very rough histogram with Poisson errors. Indeed, our estimator could find more general use if the raw data are bin counts of fixed width δ, and if δ is smaller than the optimal bin width. The calibration parameter for our second estimator is the penalty weight, λ, rather than δ.

APPENDIX A: SOLUTION DERIVATIONS

The solution of all our constrained optimization problems involve finding a stationary point of a general quadratic function $f(\mathbf{x}) = \mathbf{x}^T\mathbf{A}\mathbf{x} + \mathbf{x}^T\mathbf{b} + \mathbf{c}$, where $\mathbf{A}$ is a $p \times p$ symmetric matrix. The gradient vector, $\nabla_{\mathbf{x}} f(\mathbf{x})$, contains all p of the partial derivatives, $\partial/\partial x_k$, $k = 1, \ldots, p$. It is straightforward to show that

$$\nabla_{\mathbf{x}} f(\mathbf{x}) = 2\mathbf{A}\mathbf{x} + \mathbf{b}\,.$$

For example, $\partial/\partial x_k(\mathbf{x}^T\mathbf{b}) = \partial/\partial x_k(\sum_i x_i b_i) = b_k$. So $\nabla_{\mathbf{x}}(\mathbf{x}^T\mathbf{b}) = (b_1\, b_2\, \cdots\, b_p)^T = \mathbf{b}$.

For the first constrained problem given in Equation (5), we form the Lagrangian (with Lagrange multiplier $\boldsymbol{\gamma} \in \Re^M$ for the M constraints)

$$L(\mathbf{g}, \boldsymbol{\gamma}) = \mathbf{g}^T\mathbf{A}\mathbf{g} + \boldsymbol{\gamma}^T(\mathbf{W}\mathbf{g} - \mathbf{f})\,,$$

or, equivalently, $\mathbf{g}^T\mathbf{A}\mathbf{g} + \mathbf{g}^T\mathbf{W}^T\boldsymbol{\gamma} - \mathbf{f}^T\boldsymbol{\gamma}$, when convenient. The solution, $\hat{\mathbf{g}}$, solves the pair of vector equations

$$\nabla_{\mathbf{g}} L(\mathbf{g}, \boldsymbol{\gamma}) = 2\mathbf{A}\mathbf{g} + \mathbf{W}^T\boldsymbol{\gamma} = \mathbf{0}_m \qquad \text{(A.1)}$$

$$\nabla_{\boldsymbol{\gamma}} L(\mathbf{g}, \boldsymbol{\gamma}) = \mathbf{W}\mathbf{g} - \mathbf{f} = \mathbf{0}_M\,. \qquad \text{(A.2)}$$

If $\mathbf{A}$ is invertible, then Equation (A.1) implies

$$\hat{\mathbf{g}} = -\frac{1}{2}\mathbf{A}^{-1}\mathbf{W}^T\boldsymbol{\gamma}\,.$$

Substituting into Equation (A.2) and solving gives $\boldsymbol{\gamma} = -2(\mathbf{W}\mathbf{A}^{-1}\mathbf{W}^T)^{-1}\mathbf{f}$. Finally,

$$\hat{\mathbf{g}} = \mathbf{A}^{-1}\mathbf{W}^T(\mathbf{W}\mathbf{A}^{-1}\mathbf{W}^T)^{-1}\mathbf{f}\,.$$

Note that $\mathbf{W}\hat{\mathbf{g}} = \mathbf{W}\mathbf{A}^{-1}\mathbf{W}^T(\mathbf{W}\mathbf{A}^{-1}\mathbf{W}^T)^{-1}\mathbf{f} = \mathbf{f}$, as desired.

If the matrix $\mathbf{A}$ is not invertible, then $\hat{\mathbf{g}}$ is obtained by multiplying Equation (A.1) by $\mathbf{W}$, adding to Equation (A.2), and solving for $\boldsymbol{\gamma}$. Then

$$\hat{\mathbf{g}} = [P + 2(I - P)\mathbf{A}]^{-1}\,\mathbf{W}^T(\mathbf{W}\mathbf{W}^T)^{-1}\mathbf{f}\,,$$

where $P = \mathbf{W}^T(\mathbf{W}\mathbf{W}^T)^{-1}\mathbf{W}$.

For the second constrained problem given in Equation (6) with area constraint $\mathbf{g}^T\mathbf{1}_m - 1/\delta = 0$, the Lagrangian is (with single Lagrange multiplier γ)

$$L(\mathbf{g}, \gamma) = (\mathbf{W}\mathbf{g} - \mathbf{f})^T\mathbf{V}^{-1}(\mathbf{W}\mathbf{g} - \mathbf{f}) + \lambda \cdot \mathbf{g}^T\mathbf{A}\mathbf{g} + \gamma\,(\mathbf{g}^T\mathbf{1}_m - 1/\delta)\,.$$

The stationary point is found by solving

$$\nabla_{\mathbf{g}} L(\mathbf{g}, \gamma) = 2\mathbf{W}^T\mathbf{V}^{-1}(\mathbf{W}\mathbf{g} - \mathbf{f}) + 2\lambda\mathbf{A}\mathbf{g} + \gamma\,\mathbf{1}_m = \mathbf{0}_m \qquad \text{(A.3)}$$

$$\nabla_{\gamma} L(\mathbf{g}, \gamma) = \mathbf{1}_m^T\mathbf{g} - 1/\delta = 0\,. \qquad \text{(A.4)}$$

Multiply Equation (A.3) by $\mathbf{1}_m^T$, add to twice Equation (A.4), and solve for the Lagrange multiplier γ. Substituting back into Equation (A.3), and solving for $\mathbf{g}$ gives

$$\hat{\mathbf{g}} = \left[(I - \frac{J}{m})(\mathbf{W}^T\mathbf{V}^{-1}\mathbf{W} + \lambda\mathbf{A}) - \frac{J}{m}\right]^{-1} \times \left[(I - \frac{J}{m})\mathbf{W}^T\mathbf{V}^{-1}\mathbf{f} + \frac{\mathbf{1}_m}{m\delta}\right],$$

where J is the $m \times m$ matrix $\mathbf{1}_m\mathbf{1}_m^T$.

[Received August 2007. Revised March 2008.]

REFERENCES

Boneva, L. I., Kendall, D. G., and Stefanov, I. (1971), "Spline Transformation: Three New Diagnostic Aids for the Statistical Data-Analyst" (with discussion), *Journal of the Royal Statistical Society*, Series B, 33, 1–70.

de Boor, C. (1979), *A Practical Guide to Splines*, Berlin: Springer.

Eilers, P. H. C., and Marx, B. D. (1996), "Flexible Smoothing With *B*-Splines and Penalties," *Statistical Science*, 11, 89–102.

Freedman, D., and Diaconis, P. (1981), "On the Histogram as a Density Estimator: L_2 Theory," *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 57, 453–476.

Freedman, D., Pisani, R., and Purves, R. (1978), *Statistics*, New York: W. W. Norton & Co Inc.

Good, I. J., and Gaskins, R. A. (1980), "Density Estimation and Bump-Hunting by the Penalized Likelihood Method Exemplified by Scattering and Meteorite Data," *Journal of the American Statistical Association*, 75, 42–56.

Kooperberg, C., and Stone, C. J. (1991), "A Study of Logspline Density Estimation," *Computational Statistics and Data Analysis*, 12, 327–347.

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

Rudemo, M. (1982), "Empirical Choice of Histograms and Kernel Density Estimators," *Scandinavian Journal of Statistics*, 9, 65–78.

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

Scott, D. W., and Sagae, M. (1997), "Adaptive Density Estimation with Massive Data Sets," in *ASA Proceedings of the Statistical Computing Section*, pp. 104–108.

Scott, D. W., Tapia, R. A., and Thompson, J. R. (1980), "Nonparametric Probability Density Estimation by Discrete Maximum Penalized-Likelihood Criteria," *The Annals of Statistics*, 8, 820–832.

Scott, D. W., and Terrell, G. R. (1987), "Biased and Unbiased Cross-Validation in Density Estimation," *Journal of the American Statistical Association*, 82, 1131–1146.

Terrell, G. R. (1990), "Linear Density Estimates," in *ASA Proceedings of the Statistical Computing Section*, pp. 297–302.

Wahba, G. (1981), "Data-Based Optimal Smoothing of Orthogonal Series Density Estimates," *The Annals of Statistics*, 9, 146–156.

Wainer, H. (2006), "Finding What is not There Through the Unfortunate Binning of Results: The Mendel Effect," *Chance*, 19, 49–52.

Wand, M. P. (1997), "Data-Based Choice of Histogram Bin Width," *The American Statistician*, 51, 59–64.

This article has been cited by:

1. Lorraine Denby, Colin Mallows. 2009. Variations on the HistogramVariations on the Histogram. *Journal of Computational and Graphical Statistics* **18**:1, 21-31. [Abstract] [PDF] [PDF Plus] [Supplementary material]