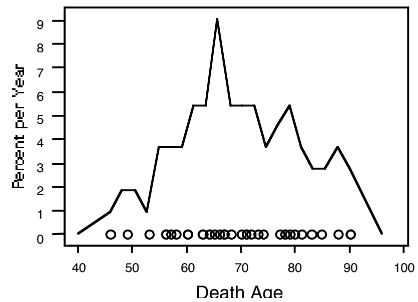
The Role of Weighted Averages in Statistics Education and Practice
K.L. Weldon
Department of Statistics and Actuarial Science
Simon Fraser University
Vancouver, Canada

Averaging data to amplify signal and reduce noise is probably the most common strategy in statistics. The common methods of estimation and hypothesis testing, including regression and analysis of variance are all firmly based in this strategy. The simplicity of the idea of averaging helps to encourage its widespread use by data analysts, even those without extensive statistical training. The more advanced techniques like density estimation, time series analysis, and nonparametric smoothing appear to involve more sophisticated strategies. However, the only slightly more complicated strategy of weighted averaging can be understood and explained by most users of statistics, and I will illustrate how these apparently advanced techniques can be portrayed as simple weighted averages. These “advanced” techniques are actually simple to understand and to explain, and with software that is now available, can be used reliably by researchers without extensive statistical backgrounds.

Most researchers contemplating even the simplest analysis of data will be familiar with the idea of averaging: $\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$. To emphasize the way in which this operation gives equal weight to all observations, we can write it $\bar{x} = \sum_{i=1}^{n} w_i x_i$ where $w_i = 1/n$ for $i = 1,2,\ldots,n$. Now it is a small step to suggest allowing other sets $\left\{ w_i : 0 < w_i < 1 \; and \; \sum_{i=1}^{n} w_i = 1 \right\}$. One way to motivate this is to present discrete data: 1,1,1,1,3,3,3,3,3,3,3,4,4 and suggest the arithmetic average could be computed as a weighted average of 1,3, and 4. Of course, even first courses in statistics will use this approach to motivate the expected value formula. Less common is the use of the weighted average to construct a variation of the common histogram, which we illustrate below:

1. Density Estimation: Consider the data provided by the death ages of American Presidents. First define an equally-spaced grid across the range of the data. Then for each grid value, scan the entire data set noting the distance of each data point from the grid point. For some arbitrary distance d (such as a tenth of the range), count up the number of data values within d of the grid point, and plot that value at the grid point. Note this is a weighted sum where the weights are 1 or 0. Do this for each grid point and then plot the result. Except for the vertical scale, the resulting graph would look like:

Fig. 2 Windowgram



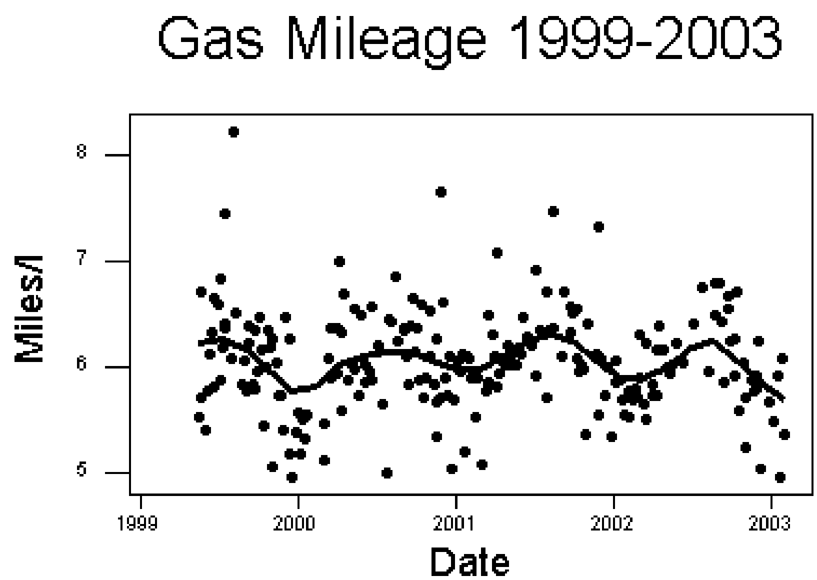
This crude kernel density estimator can be scaled to percent per year as shown, if desired. The obvious relationship between a weighted sum and a weighted average can be explained when the scaling is important. Also, a good default value for the half-width of the kernel is $Range/\sqrt{n}$.

However neither the scaling nor the default value of d need be understood to make use of this alternative to the traditional histogram. Trial and error can be used to choose a reasonable value of d – even the naïve user will realize that some local deviations are likely due to random variation.

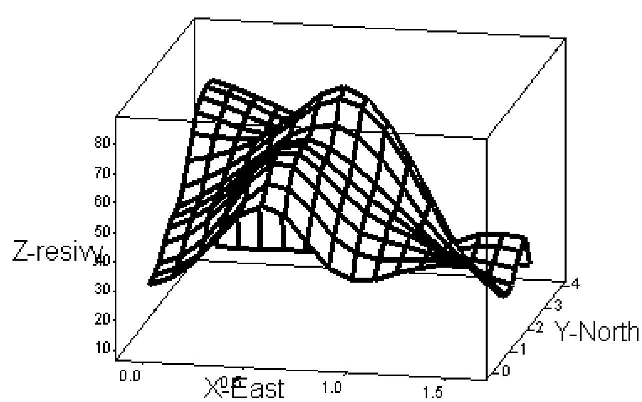
What has been shown is that a density estimator that is really a weighted sum of frequencies of data values can be motivated by a simple count over intervals of data. To modify this procedure by using a more gradually decreasing kernel, instead of having weights decrease suddenly from 1 for close points to 0 for points more distant than d from a grid point, seems another small step, and would produce smoother density estimates. The weighted sum of frequencies, with weights determined by distance from a grid point, gives rise to a flexible class of simple density estimators. To put it another way, the histogram can be improved upon by using a weighted average strategy.

2. Time Series Analysis: The weighted average process can be used in time series analysis as well: both autoregressive and moving average operators are clearly weighted averages. However there is a less obvious way that weighted averages can help in the time series toolkit. Many time series will contain seasonal or other patterns. When the variability around the trend swamps the trend itself, some smoothing is required to extract the trend. There are many techniques for doing this, but one of the simplest to describe uses a weighted average approach. At each grid value along the time axis, simply compute a weighted average of the nearby points. The following graph is computed this way and the seasonal trend in this gas consumption data shows through the considerable noise very clearly. The weighting function in this case was $\exp(-cd^2)$ where d is the distance to the grid point, and c is a constant determined either by a default based on the x-range of the data, or by trial and error.

This nonparametric smoothing method is not limited to time series. For any data in which the objective is to fit a curve y=f(x) to predict y from x, the same weighted average smoother is appropriate. The avoidance of guessing the parametric form of the fit is a great benefit. Residual analysis is still possible based on this nonparametric fit.



3. Nonparametric Smoothing: As already mentioned, the weighted average smoother described for the time series example above can also be used for estimation of functions $y=f(x)$ based on any bivariate data set. However, suppose we have a trivariate data set (y,x_1,x_2). The same approach is easily extended. We use a two-dimensional grid in the x_1-x_2 plane, and again compute weighted averages of nearby values. The result can be displayed as a contour plot or as a wireframe plot.



The data portrayed here is taken from a data set of 1000 soil resistivity measurements over a rectangular agricultural area, taken from Cleveland(1993).

Discussion:

The techniques described here are motivated by the 1993 book Visualizing Data by W.S. Cleveland. He suggests more sophisticated methods of nonparametric smoothing in two and more dimensions. However, a student with a good understanding of the simple techniques described here could grasp Cleveland's methods at the next stage. Meanwhile, the raw methods used here would be adequate for many applications, and perhaps easier to explain to end-users.

REFERENCES

Cleveland, W.S. (1993) *Visualizing Data*. Hobart Press, Summit, NJ.
Freedman, D., Pisani, R. ,Purves, R. (1998) *Statistics.* Third Edition. Norton. New York
Moore, D.S. and McCabe, G.P. (1999) *Introduction to the Practice of Statistics.* Third Edition. Freeman. NY.
Wild, C.J and Seber, G.A. (2000) *Chance Encounters: A first Course in Data Analysis and Inference.* Wiley. New York.