Weighted Averages

Abstract A recent discussion on CASNET raised the issue of what level of mathematics actuaries typically use. Yet weighted averages, which are a constant element of most actuarial work, received scant attention. And in fact it is difficult to find a CAS paper on this universally pervading topic. The present work attempts to fill in that gap. The key issue is not how to compute a weighted average but what weights to use.

Arithmetic Weighted Averages

Typically weighted averages are taken of ratios, such as loss ratios. An example would be weighting loss ratios by territory to get a statewide loss ratio. For instance, denote the ratios by $loss_j/premium_j$. Weighting these by premium would use the weights: $premium_j/\sum_j premium_j$. Doing the algebra shows that the resulting weighted average would be $[\sum_j loss_j]/[\sum_j premium_j]$, which is the overall statewide average loss ratio.

This is an illustration of the general rule for arithmetic weighted averages of ratios: weight by the denominators. As simple as this seems, it is sometimes ignored in practice and even more often is arrived upon anew for a single application after lengthy and tedious debate about what weights to use. It is not unusual, however, for actuaries to be dealing with ratios of ratios, in which case it is not always clear what the denominator should be for application of this rule. In such cases it is useful to go back and do the algebra to uncover the meaning of the weighted average, as in the above case of loss ratios, where the result was the statewide ratio.

When should this rule be applied? Basically when the purpose of the weighted average is to produce the ratio being weighted at a more aggregate level, which is typical of actuarial situations. However there are other applications for which this method would not be optimal, such as weighting for the purpose of estimating the mean ratio for a random draw out of a population of ratios. Statistical methods for such problems are discussed later.

Weights Inverse to Variance

A general rule of thumb in statistics is to weight different observations inversely to the variance of each. This can arise, for example, when different samples are taken of the same random process, and the samples are to be combined to get an overall estimate. If the observations are independent, this method leads to the minimum variance estimate of the quantity being estimated.

If X and Y are independent, $Var[aX + (1-a)Y] = a^2VarX + (1-a)^2VarY$. This takes a minimum where its derivative is zero, which gives a(VarX + VarY) = VarY. Divide both sides by VarXVarY and solve for a to get:

a = [1/VarX]/[1/VarX + 1/VarY].

This could be used to combine several approximate estimates of a physical process, like the speed of light through lithium, each with its own variance. An actuarial example is age-to-age factors. The variance of each factor at a given age might be inversely proportional to the cumulative losses emerged by the previous age, or it might be unaffected by loss volume. The latter would imply using a straight average of the factors, while the former would weight by losses in the denominator of the factor. As in the arithmetic case, this would give the factor calculated by the aggregate of losses at one age divided by the aggregate of losses at the previous age. Yet if the variance of the individual factors is in fact unaffected by loss volume, the straight average would give a better estimate. This is sometimes a subtle effect to measure, however.

As an example, ten years of age 0 to age 1 loss development factors were simulated two ways for a growing book of business. Factors were drawn at random with a mean of 1.3. The variance used for each draw in one case was constant, and in the other case was inversely proportional to the losses at age 0. The actual simulation procedure for the deviation from the mean for the factor for a given accident year was to draw a standard normal deviate and multiply by the standard deviation for that accident year. As the standard deviation in the second case was inversely proportional to the square root of the age 0 losses, the deviations were produced by multiplying a random factor by one over the square root of those losses. Thus the absolute value of the deviations should grow linearly (but with a random factor) with the inverse of the square root of the age 0 losses. This should not happen with in the constant variance case.

The question is, will it be possible to differentiate these cases by looking at the factors. Table 1 shows the standard deviations used and the resulting absolute deviations for each case for a single simulation. The average standard deviation of 0.081 was used for the constant case. Also shown is a linear fit of the absolute deviations to one over the square root of the loss at age 0.

Table 1

Loss @ 0	**Standard Deviation Variable Case**	**Absolute Deviation Variable Case**	**Fit**	**Absolute Deviation Constant Case**	**Fit**
100,000,000	0.100	0.021	0.110	0.078	0.002

110,000,000	0.095	0.065	0.100	0.037	0.015
121,000,000	0.091	0.077	0.090	0.058	0.028
133,100,000	0.087	0.115	0.081	0.0248	0.040
146,410,000	0.083	0.157	0.072	0.0153	0.051
161,051,000	0.079	0.009	0.064	0.040	0.062
177,156,100	0.075	0.069	0.056	0.019	0.072
194,871,710	0.072	0.030	0.048	0.089	0.082
214,358,881	0.068	0.014	0.041	0.184	0.091
235,794,769	0.065	0.049	0.034	0.074	0.100
	0.081				

In the variable variance case, the absolute deviations do decrease as losses increase, as shown by the fitted line. This does not happen in the constant variance case – in fact it looks like there is an upward trend, even though these were random draws from a level process. In the variable variance case, the fitted slope coefficient is 2196 with a standard deviation of 1543, which is at best marginally significant. This was typical of other simulations as well. By looking at trends in the variability of the factors, sometimes you could tell the constant variance and variable variance cases apart and sometimes you could not. Graphs of these two cases are in Figures 1 and 2. Constant variance as the simpler assumption would seem to be the appropriate default.

Figure 1

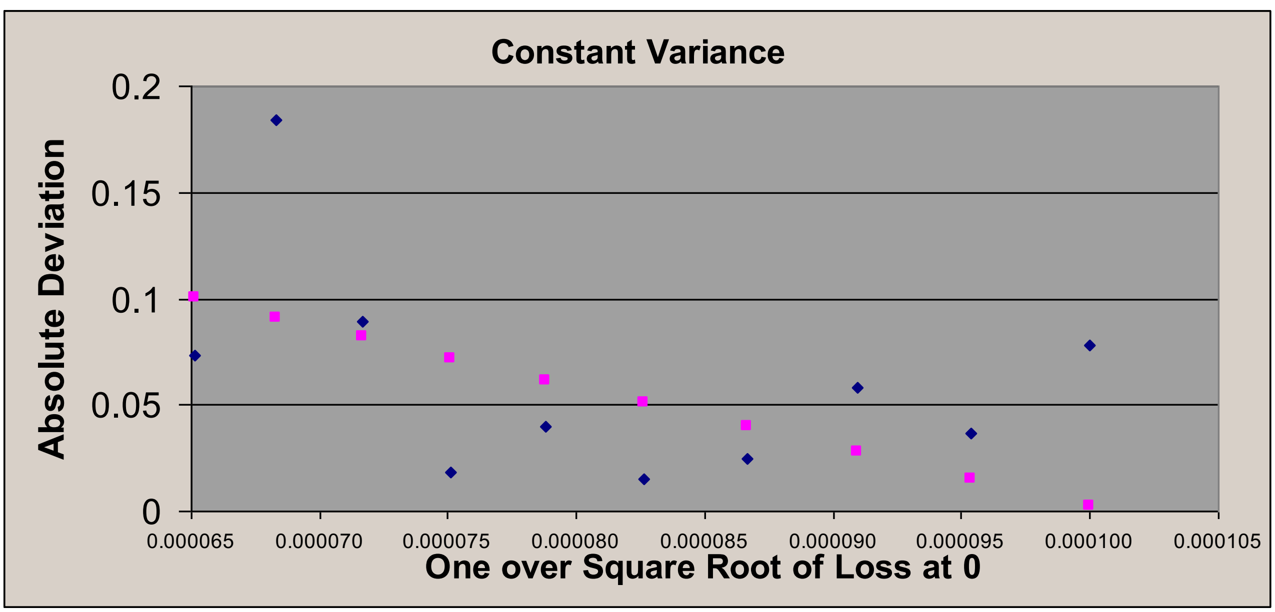
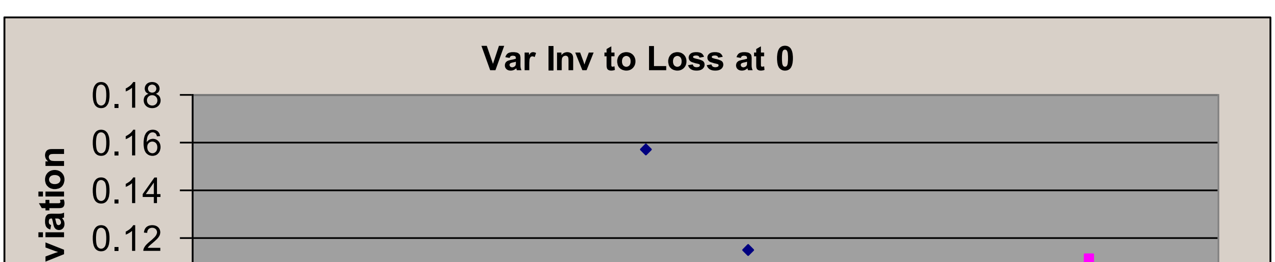


Figure 2



Moving Averages

These are weighted averages of a time series. A typical actuarial application is to give declining weights to older accident year factors in a development triangle. For example, some actuaries use weights of 5,4,3,2,1 for the five most recent years. This might be appropriate, for instance, if the development process were evolving in some fashion that made the older years less relevant.

However, techniques like this should not be used in a rote fashion. If the development process has been stable, giving less weight to the older years reduces the accuracy of the estimate. Either looking to see if there is a trend in the factors, or looking at the variance of the recent factors and testing the older ones to see if they are significantly different can test the stability of the factors.

One method of arriving at the set of declining factors to use is exponential smoothing. For the jth most recent factor, this uses the weight exp(-jx) for some fixed x. This results in weights proportional to 1, y, y^2,y^3, . . ., where y = exp(-x). For instance, for y=0.7 the first few weights are 1.000, 0.700, 0.490, 0.343, 0.240, 0.168. This illustrates one aspect of exponential smoothing – the drop-off in the weights slows down as you go along.

Centered Moving Averages

Smoothing of economic time series is often done to try to separate random fluctuations from an underlying systematic component. An insurance application might be to identify a smooth underwriting cycle underlying the random fluctuations of loss ratios. Centered moving averages can be tried for this. For instance, a five point centered moving average gives equal weight to an observation and the two observations on either side of it.

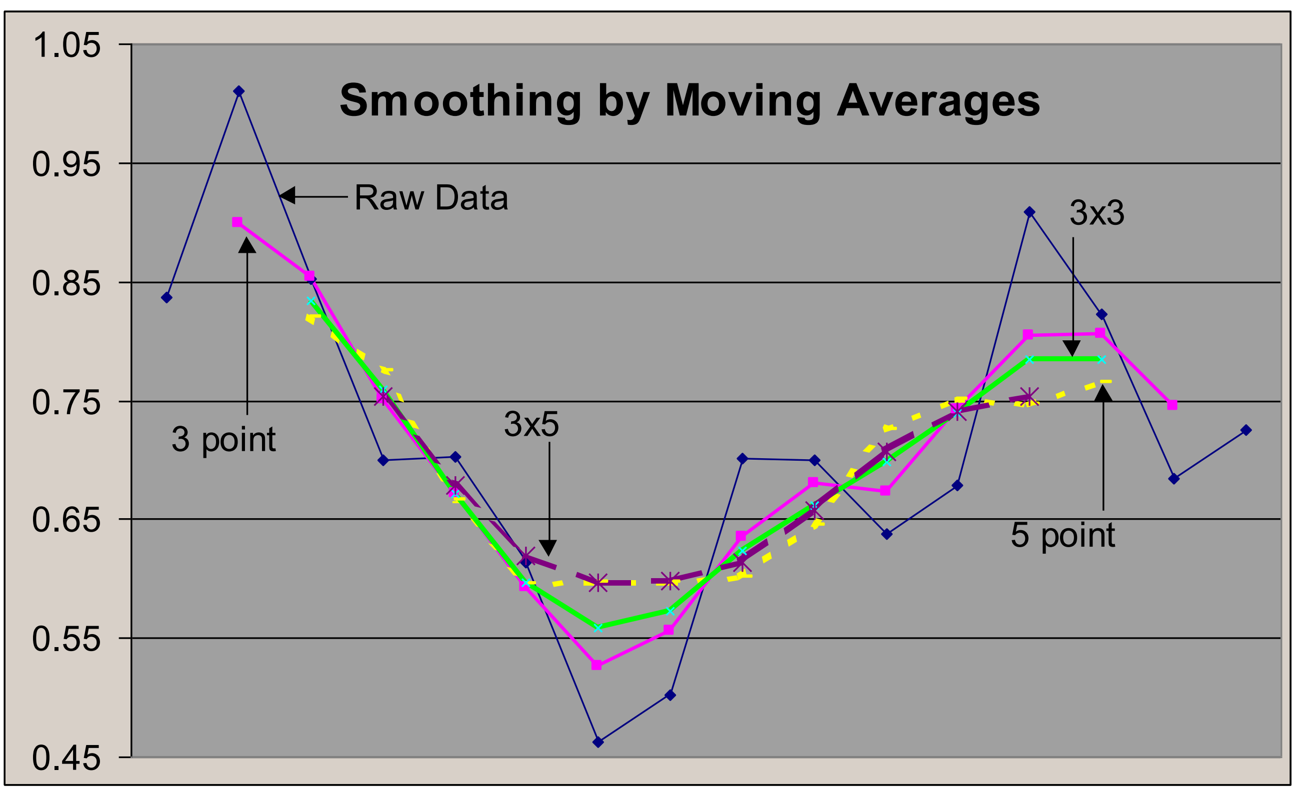
Sometimes taking moving averages of moving averages creates more smoothing. A 3x5 centered moving average takes a three point centered average of a five point centered average. This produces weights proportional to 1,2,3,3,3,2,1 for the seven points centered at each observation.

There is an endpoint problem with centered averages, however. The 3x5 average, for example, cannot be calculated for the three endpoints at either end of the series. There is no established way for treating this. If there are enough points, the endpoints can sometimes just be ignored. Alternatively, some analysts reverse the series at both ends. E.g., Ben Zehnwirth once suggested something like this. For the 3x5 average the weights for the last six points that produce the last three weighted points are respectively [1,2,3,3,3,3], [0,1,2,3,4,5], and [0,0,1,3,5,6]. This often appears to give reasonable smoothing but it should not be interpreted as indicating a new trend or a slowing down of an old trend. Also, the smoothed endpoints will change when new observations become available.

Centered moving averages can also use exponential smoothing in both directions to compute the weights. Unless the series is very long, some technique like reversing the series at the endpoints will be needed to avoid dropping too many observations.

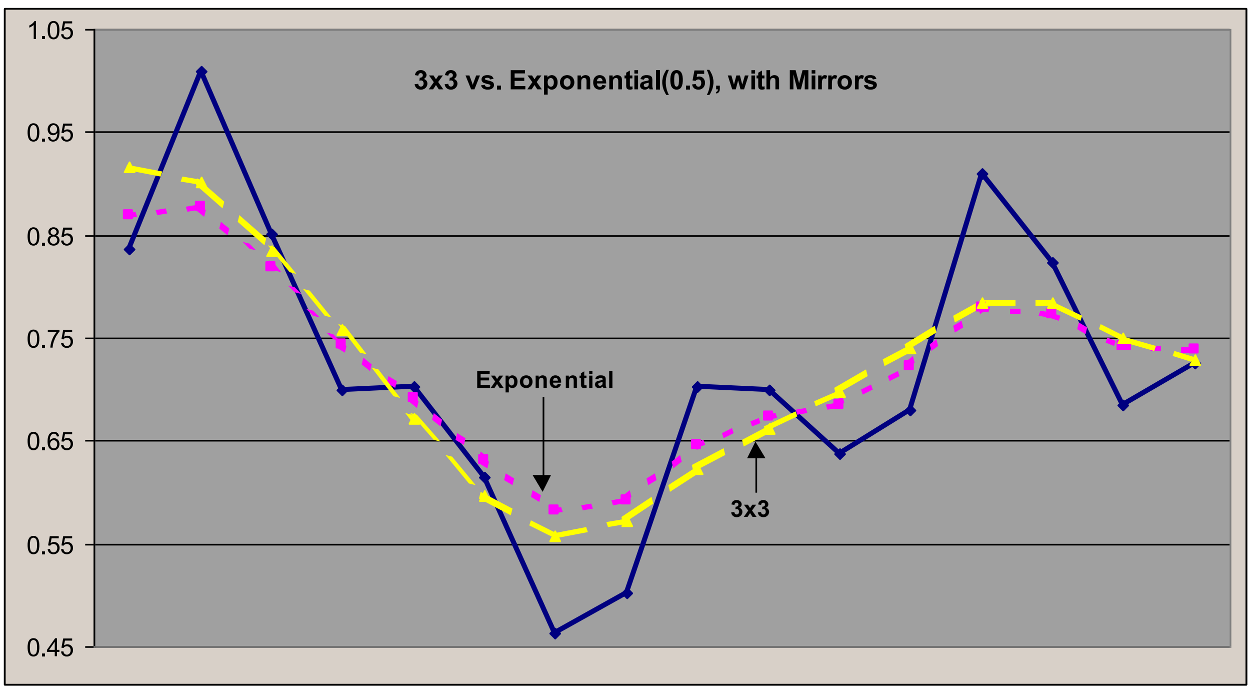
As an illustration, loss ratios were simulated for a hypothetical line of insurance for a sixteen year period, which reflected in this case somewhat more than one underwriting cycle. Figure 3 illustrates the effect of smoothing by centered moving averages.

Figure 3



The three point centered average loses only one point at each end, and it is the closest to the raw data, but it seems a little over-responsive to wiggles in the data – i.e., it does not smooth enough. The five point average and the 3x3 average each lose two endpoints. The 3x3 average is closer to the original data overall, and appears a little smoother as well. The 3x5 average loses another point at each end, is further from the data, and is not really noticeably smoother than the 3x3 average, which thus looks like the best of these approaches for this particular data.

Next, a test was made of the idea of reflecting a series at each endpoint to continue the averaging even for the endpoints themselves. This was done for the same data and the 3x3 centered moving average and an average using exponentially declining weights in each direction. The latter gave a weight of 1/3 to the centered data value and then reduced the weight by applying a factor of 0.5 at each subsequent step away from the center. This procedure is actually determined by choosing the factor of 0.5. The initial weight of 1/3 is what is needed for the weights to sum to unity. This can be determined for any weighting factor w by summing the proportional weights $1+2(w+w^2+w^3+...)$, up



to whatever point the series gets truncated, to get a normalizing constant. When enough points are used, the central weight should approach (1-w)/(1+w). The smoothed series for this example are shown in Figure 4.

Figure 4

With this weighting, the exponential smoothing series looks a little less responsive at the biggest dip and somewhat more responsive at the smaller swings than does the 3x3 average. It probably is not an improvement for this data. The smoothing at the endpoints looks plausible for both methods.

Credibility

Credibility theory is the most purely actuarial approach to weighted averages. Consider the Buhlmann-Straub case, where a typical example is to weight classification pure premiums against the statewide average. Stuart Klugman once showed me that this is an example of the statistical inverse variance weight rule of thumb.

The setup is that the jth class has a mean pure premium m_j, and these are distributed with mean m and variance t^2. A unit of loss has variance s^2, so a class with p independent exposure units has loss variance ps^2 and pure premium variance s^2/p. There are two possible estimates that could be used for the class pure premium – the statewide mean m, which has variance t^2, or the class data, with variance s^2/p. If each of these estimates is weighted inversely to its variance, the class data gets a weight of $[p/s^2]/[p/s^2+1/t^2]$. A little algebra gives the usual credibility formula for the weight of $p/[p+k]$, with $k=s^2/t^2$.

An often overlooked subtlety of this approach is that m is not the overall state pure premium, which is greatly influenced by the high-exposure classes. Every class pure premium must be considered a random draw with the same variance in this setup. In a state with one large class and a few small ones, the large class would be closer to the overall statewide pure premium than the small ones would be, so it would not have the same variance from the overall mean. To get the mean of all the individual classes you have to do a specific weighted average of the individual class means. The weights derived for this in credibility theory are just the individual class credibilities.

This answer also falls out directly using the inverse variance rule. Each class pure premium varies from the mean with variance t^2 and further by its estimation noise which has variance s^2/p. Thus a class with exposure p should get a weight proportional to $1/[t^2 + s^2/p]$. Multiplying numerator and denominator by p/t^2 shows that the weight has to be proportional to $[p/t^2]/[p+k]$. The $1/t^2$ in the numerator can be ignored, as it will be a constant for each class, so the weight will be proportional to the credibility $p/[p+k]$ for the class.

Of course once we are estimating the statewide mean m, it has some estimation error, so the variance of class means from the estimated m is higher than t^2. This changes the credibilities to some degree. In practice the variances s^2 and t^2 are usually estimated as well, which changes the situation even further. The statistical approach known as empirical Bayes addresses this to some degree, and they end up with an adjustment factor to the credibility weight. This adjustment however ignores the fact that credibility factors are usually capped to stay in the range [0,1]. In some cases this capping alone replaces the need for the adjustment factor. My chapter on credibility in the CAS textbook addresses this in more detail.

Empirical Bayes work has sometimes been able to use some mysterious mathematical magic (arcsine transforms of binomial variates?) to avoid the need for estimating variances. The full classification credibility problem including estimation of all the means and variances has been attacked (Klugman) using Bayesian methods, but it gets complicated. The simple formula for the weighted average does not hold up under that level of scrutiny.

Some other generalizations are in the CAS textbook. One such is to recognize that class properties change over time, so the estimation of the mean from class data has some element of variance that does not decrease with class size. Another, when applied to individual risk rating for commercial risks, is to see if the large risks are more homogeneous in loss experience than are the smaller risks. This could happen from some government supervisory approaches, for example. If it is that way, further adjustments to the credibility formula are needed.

Credibility theory appears to be non-parametric – i.e., the formulas do not seem to depend on what distributions are followed. This is an illusion, however. The optimization method selected – linear least squares – is really optimal only for normal and some closely related distributions. The credibility chapter shows that with heavier-tailed distributions it is better to do the credibility estimation in the logs of the data, exponentiate the results and then rebalance for the bias produced by averaging logs. The Bayesian work by Klugman uses normal distributions, which is not a limitation if the problem is a reasonable fit with the least squares paradigm. Some of the apparently strange devices used in historical actuarial methods, like full credibility for 50 claims for serious workers compensation losses, make sense as linear approximations to an often highly non-linear problem.

The intricacy of the mathematics builds up if you want to take all of these issues into account, but it may make commercial sense to do so as class-plan based competition becomes increasingly computer intensive.

Credibility for Moving Averages

Gerber and Jones worked out an interesting case. Suppose you need to estimate the next observation of a time series that has a mean changing over time by a given variance, but all you have is a single observation at each period, drawn with another variance. An example might be the statewide loss ratio for a line of business. The credibility estimate then is given by a weighted average of the observations where