Fitting Moments with Weights

Daniel R. Corro

Fitting Moments with Weights

Dan Corro
National Council on Compensation Insurance, Inc.
June, 2002

Abstract: *This note investigates ways to fit individual claim loss data to a prior known "underlying severity level" by adjusting the relative importance, or weight, assigned to each claim. Here, "underlying severity level" is measured by the weighted mean cost per case. The paper also generalizes the approach to accommodate fitting higher moments of the loss distribution, especially the variance. It establishes the existence of an optimal reweighting, but whose calculation may be too difficult for practical application. To address this, the paper describes two easier calculations, one designed to fit only the mean and another to fit both mean and variance.*

Section I: Setup and Notation

Let X be any finite set, by a *weight on* X we simply mean a non-negative real-valued function $\omega : X \to [0,\infty)$. In this case will also refer to ω as a *weight* and refer to the pair (X,ω) is a *weighted set.* But we will often abuse this formality and just refer to X as weighted by ω. For any finite set X, we let $|X|$=number of elements in X. When X is weighted by ω, we use the notation:

$$|A|_\omega = \sum_{x\in A} \omega(x), \text{ for any subset } A \subseteq X.$$

We note two simple properties that a weight ω on X may or may not have:

ω is *positive* if and only if $\omega(x) > 0$ for every $x \in X$

ω is a *probability weight* if and only if $|X|_\omega = 1$.

It is clear that the concept of a discrete probability density on X exactly coincides with what we are here calling a probability weight.

Now let $X \subset \mathbb{R}$ be any finite set of real numbers and ω a weight on X. By combining the weights of elements of X that are equal, we can without any loss of generality write $X = \{x_1 < x_2 < ... < x_n\}$ as a series of n distinct numbers in ascending order. Think of the x_i as representing the distinct loss amounts from the claim sample X, arranged in increasing order to facilitate a size of loss analysis. Now take any $z \in \mathbb{R}$ with $x_1 < z < x_n$. It is intuitively clear that there exits a weight υ on X for which z is the weighted mean:

$$z = \frac{\sum_{x \in X} \upsilon(x)x}{\sum_{x \in X} \upsilon(x)} = \frac{\sum_{x \in X} \upsilon(x)x}{|X|_\upsilon} = \mu_\upsilon(X).$$

If we define yet a third weight ρ on X by setting $\rho = \dfrac{\upsilon |X|_\omega}{\omega |X|_\upsilon}$. Then we can think of ρ as a multiplicative adjustment factor to the weight ω that reweights the weighted set X to give it the given mean z while holding the total weight constant.

Section II: Moments of Finite Claim Samples

This paper pursues the question of how to come up with an appropriate υ. For this purpose, we introduce the *formal moments* of X, relative to any *function* $\upsilon : X \to \mathbb{R}$

$$\mu_k = \mu_k(X, \upsilon) = \frac{\sum_{x \in X} \upsilon(x) x^k}{|X|_\upsilon}, \quad 0 \le k \le n-1.$$

Observe that when υ is a probability weight, this is just the usual first $n-1$ moments of the claim sample X. It turns out that for any vector of potential formal moments of X, say $m = (1, m_1, .. m_{n-1})$, there is a uniquely defined function $\upsilon(m) : X \to \mathbb{R}$ such that:

(*) $$m_k = \mu_k(X, \upsilon), \quad 0 \le k \le n-1.$$

To verify this, recall the $n \times n$ Van der Monde matrix:

$$V = V(X) = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_n \\ \vdots & \vdots & \vdots & \vdots \\ x_1^{n-1} & x_2^{n-1} & \cdots & x_n^{n-1} \end{bmatrix}.$$

whose determinant:

$$Det(V) = \prod_{1 \le j < i \le n} (x_i - x_j) > 0.$$

provides a standard exercise in introductory linear algebra textbooks. The verification is by induction on n. Case $n = 1$ holds vacuously and case $n = 2$ is clear. Regard the x_i as constants and construct the $n-1$ degree polynomial:

$$p(y) = Det\begin{bmatrix} 1 & 1 & \cdots & 1 & 1 \\ x_1 & x_2 & \cdots & x_{n-1} & y \\ \vdots & \vdots & & \vdots & \vdots \\ x_1^{n-1} & x_2^{n-1} & \cdots & x_{n-1}^{n-1} & y^{n-1} \end{bmatrix}.$$

Note that substituting y by any of $x_1, \ldots, x_{n-1}$ results in a matrix with two identical columns. But then clearly $p(y)$ has the distinct roots $x_1, \ldots, x_{n-1}$, and we may write $p(y) = a \prod_{1 \le i < n} (y - x_i)$, where the constant a is the coefficient of y^{n-1}. But expanding the determinant along column n and invoking the induction hypothesis:

$$a = Det\begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_{n-1} \\ \vdots & \vdots & \vdots & \vdots \\ x_1^{n-2} & x_2^{n-2} & \cdots & x_{n-1}^{n-2} \end{bmatrix} = \prod_{1 \le j < i \le n-1} (x_i - x_j).$$

Whence:

$$Det(V) = p(x_n) = a\prod_{1\le i<n}(x_n - x_i) = \prod_{1\le j<i\le n-1}(x_i - x_j)\prod_{1\le i<n}(x_n - x_i) = \prod_{1\le j<i\le n}(x_i - x_j),$$

that completes the induction.

Now we can naturally identify any function $\upsilon : X \to \mathrm{R}$ with the row vector $(\upsilon(x_1), \upsilon(x_2), \ldots \upsilon(x_n))$. With this notation, observe that (*) is just the matrix equation: $V\upsilon^T = m^T$. Since the matrix V is nonsingular, the function $\upsilon : X \to \mathrm{R}$ can be calculated from $\upsilon^T = V^{-1}m^T$, establishing both existence and uniqueness of υ. In theory, this provides a way of determining whether a weight υ exists on X that reweights the claims to fit the given set of n moments, and even provides a way to calculate it. In practice, however, the claim sample may be very large and this may not be very practical.

More likely, we are only concerned with fitting the first few moments of the claim sample X to a set of moment values derived from empirical data, say $\hat{m} = (\hat{m}_0 = 1, \hat{m}_1, \ldots \hat{m}_k), k \le n$. The moments must be reasonable in relation to X, for example we clearly must have:

$$x_1 m_j \le m_{j+1} \le x_n m_j \; 1 \le j \le k.$$

Which would be assured, say, if all the empirical claim costs fell within the range of X.

When $k = 1$ it is clear that the set of "possible" moments over all probability weights on X is just:

$$M_1(X) = \{(1, m_1) \mid x_1 \le m_1 \le x_n\}.$$

The case $k = 2$, which corresponds to fitting both the mean and standard deviation, is more complicated and so we consider sub-cases.

Sub-case $k = n = 2$, here the reader can easily verify that:

$$M_2(X) = \{(1, m_1, m_2) \mid x_1 \le m_1 \le x_2,\ m_2 = x_2(m_1 - x_1) + m_1 x_1\}.$$

Sub-case $k = 2, n = 3$, here we claim that

$$M_2(X) = \left\{ (1, m_1, m_3) \middle| \begin{array}{c} x_1 \le m_1 \le x_3 \\ Max(x_2(m_1 - x_1) + m_1 x_1, x_3(m_1 - x_2) + m_1 x_2) \le m_2 \le x_3(m_1 - x_1) + m_1 x_1 \end{array} \right\}$$

To verify this, consider the set of 2 simultaneous equations:

$$\begin{aligned} m_1 &= \omega_1 x_1 + \omega_2 x_2 + (1 - \omega_1 - \omega_2) x_3 \\ m_2 &= \omega_1 x_1^2 + \omega_2 x_2^2 + (1 - \omega_1 - \omega_2) x_3^2 \end{aligned}$$

which may be rewritten as:

$$\begin{aligned} x_3 - m_1 &= \omega_1 (x_3 - x_1) + \omega_2 (x_3 - x_2) \\ x_3^2 - m_2 &= \omega_1 (x_3^2 - x_1^2) + \omega_2 (x_3^2 - x_2^2) \end{aligned}$$

Considering ω_1, ω_2 as unknowns, we know from the above that there is a unique solution to these equations. In fact, we let the reader verify that the solution is:

$$\omega_1 = \frac{m_2 - x_3^2 + (x_3 - m_1)(x_3 + x_2)}{(x_3 - x_1)(x_2 - x_1)}, \quad \omega_2 = \frac{x_3^2 - m_2 - (x_3 - m_1)(x_3 + x_1)}{(x_3 - x_2)(x_2 - x_1)}$$

Note too that

$$\omega_1 + \omega_2 = \frac{x_3^2 - m_2 - (x_3 - m_1)(x_2 + x_1)}{(x_3 - x_2)(x_3 - x_1)}$$

Considering ω_1, ω_2 as weights, we see that they define a probability density with moment vector $(1, m_1, m_2) \in M_2(X)$ exactly when $\omega_1 \geq 0, \omega_2 \geq 0$ and $\omega_1 + \omega_1 \leq 1$. Now the reader can easily check that:

$$\begin{aligned} x_2(m_1 - x_1) + xx_1 \leq m_2 &\quad\Leftrightarrow\quad \omega_1 + \omega_1 \leq 1 \\ x_3(m_1 - x_2) + xx_2 \leq m_2 &\quad\Leftrightarrow\quad \omega_1 \geq 0 \\ m_2 \leq x_3(m_1 - x_1) + m_1 x_1 &\quad\Leftrightarrow\quad \omega_2 \geq 0 \end{aligned}$$

from which our claim follows. There remains:

Sub-case $k = 2, n \geq 4$,

$$M_2(X) = \{(1, m_1, m_2) \mid x_1 \leq m_1 \leq x_n,\ x_n(m_1 - x_{n-1}) + m_1 x_{n-1} \leq m_2 \leq x_n(m_1 - x_1) + m_1 x_1\}.$$

To prove this, let $(1, m_1, m_2) \in M_2(X)$ and let $\omega(x_i) = \omega_i$ be the corresponding probability weight. As before, we consider the set of two simultaneous equations:

$$\begin{aligned} x_n - m_1 &= \omega_1(x_n - x_1) + \omega_2(x_n - x_2) + \ldots + \omega_{n-1}(x_n - x_{n-1}) \\ x_n^2 - m_2 &= \omega_1(x_n^2 - x_1^2) + \omega_2(x_n^2 - x_2^2) + \ldots + \omega_{n-1}(x_n^2 - x_{n-1}^2). \end{aligned}$$

Eliminating the "unknown" ω_{n-1} gives:

$$\begin{aligned} x_n^2 - m_2 - (x_n + x_{n-1})(x_n - m_1) = {} & \omega_1\left(x_n^2 - x_1^2 - (x_n + x_{n-1})(x_n - x_1)\right) \\ & + \omega_2\left(x_n^2 - x_2^2 - (x_n + x_{n-1})(x_n - x_2)\right) \\ & \quad\vdots \\ & + \omega_{n-2}\left(x_n^2 - x_{n-2}^2 - (x_n + x_{n-1})(x_n - x_{n-2})\right) \end{aligned}$$

This can be rewritten as

$$\begin{aligned} m_2 = x_n(m_1 - x_{n-1}) + m_1 x_{n-1} \quad &+ \omega_1(x_{n-1} - x_1)(x_n - x_1) \\ &+ \omega_2(x_{n-1} - x_2)(x_n - x_2) \\ &\quad\vdots \\ &+ \omega_{n-2}(x_{n-1} - x_{n-2})(x_n - x_{n-2}) \end{aligned}$$

Since the probability weights $\omega_i \geq 0$, this clearly implies that:

$$m_2 \geq x_n(m_1 - x_{n-1}) + m_1 x_{n-1}$$

Observe too that

$$(x_{n-1} - x_1)(x_n - x_1) > (x_{n-1} - x_2)(x_n - x_2) > \cdots > (x_{n-1} - x_{n-2})(x_n - x_{n-2}) > 0.$$

It follows that m_2 is maximized by assigning as much weight as possible to x_1, i.e. by making ω_1 as big as possible. Now, for fixed weighted mean m_1, the minimum x_1 gets maximum weight when it is required to offset all by itself the maximum x_n. Note that in that event:

$$m_1 = \hat{\omega}_1 x_1 + (1 - \hat{\omega}_1) x_n \quad \Rightarrow \hat{\omega}_1 = \frac{x_n - m_1}{x_n - x_1}$$

From this, we see that:

$$\begin{aligned} m_2 &\leq x_n(m_1 - x_{n-1}) + m_1 x_{n-1} + \hat{\omega}_1(x_{n-1} - x_1)(x_n - x_1) \\ &= x_n(m_1 - x_{n-1}) + m_1 x_{n-1} + (x_{n-1} - x_1)(x_n - m_1) \\ &= x_n(m_1 - x_1) + m_1 x_1 \end{aligned}$$

And we have shown that

$$M_2(X) \subseteq \{(1, m_1, m_2) \mid x_1 \leq m_1 \leq x_n,\ x_n(m_1 - x_{n-1}) + m_1 x_{n-1} \leq m_2 \leq x_n(m_1 - x_1) + m_1 x_1\}.$$

Conversely, let $(1, m_1, m_2)$ belong to the right hand side and let

$$z_1 = x_1 < z_2 = x_{n-1} < z_3 = x_n.$$

Then we find that

$$\begin{aligned}
&z_2(m_1 - z_1) + m_1 z_1 = x_{n-1}(m_1 - x_1) + m_1 x_1 \\
&\quad = x_1(m_1 - x_{n-1}) + m_1 x_{n-1} \le x_n(m_1 - x_{n-1}) + m_1 x_{n-1} = z_3(m_1 - z_2) + m_1 z_2 \\
&\quad \le m_2 \\
&\quad \le x_n(m_1 - x_1) + m_1 x_1 = z_3(m_1 - z_1) + m_1 z_1.
\end{aligned}$$

It then follows from the case $n = 3$ that $(1, m_1, m_2) \in M_2(X)$ whence:

$$\{(1, m_1, m_2) \mid x_1 \le m_1 \le x_n,\ x_n(m_1 - x_{n-1}) + m_1 x_{n-1} \le m_2 \le x_n(m_1 - x_1) + m_1 x_1\} \subseteq M_2(X)$$

and the proof is complete for the sub-case $k = 2, n \ge 4$.

That argument readily extends to:

Case $3 \le k \le n$:

$$M_k(X) \subseteq \left\{ (1, m_1, \ldots, m_k) \left| \begin{array}{c} x_1 \le m_1 \le x_n \\ x_n^j - (x_n - m_1)\left(\dfrac{x_n^j - x_{n-1}^j}{x_n - x_{n-1}}\right) \le m_j \le x_n^j - (x_n - m_1)\left(\dfrac{x_n^j - x_1^j}{x_n - x_1}\right),\ 2 \le j \le k \end{array} \right. \right\}$$

To prove this, again let $(1, m_1, m_2, \ldots, m_k) \in M_k(X)$ and $\omega(x_i) = \omega_i$ be the corresponding probability weight. We have a set of k simultaneous equations:

$$\begin{aligned}
x_n - m_1 &= \omega_1(x_n - x_1) + \omega_2(x_n - x_2) + \ldots + \omega_{n-1}(x_n - x_{n-1}) \\
x_n^2 - m_2 &= \omega_1(x_n^2 - x_1^2) + \omega_2(x_n^2 - x_2^2) + \ldots + \omega_{n-1}(x_n^2 - x_{n-1}^2) \\
&\ \ \vdots \\
x_n^k - m_k &= \omega_1(x_n^k - x^k) + \omega_2(x_n^k - x_2^k) + \ldots + \omega_{n-1}(x_n^k - x_{n-1}^k)
\end{aligned}$$

Now fix $j, 1 \le j \le n$ and define the function

$$f(x) = x_n^j - x^j - (x_n - x)\frac{x_n^j - x_{n-1}^j}{x_n - x_{n-1}}.$$

Letting $\rho = \frac{x_n}{x_{n-1}} > 1$, we see that for $x_1 < x < x_{n-1}$:

$$\frac{df}{dx} = -jx^{j-1} + \frac{x_n^j - x_{n-1}^j}{x_n - x_{n-1}} = -jx^{j-1} + \frac{x_{n-1}^j}{x_{n-1}}\frac{\rho^j - 1}{\rho - 1} = \left(1 + \rho + \ldots + \rho^{j-1}\right)x_{n-1}^{j-1} - jx^{j-1} > 0.$$

And so $f(x)$ is an increasing function on (x_1, x_{n-1}). Since $f(x_{n-1}) = 0$, we see that

$$0 < -f(x_{n-2}) < -f(x_{n-3}) < \ldots < -f(x_2) < -f(x_1).$$

Eliminating the "unknown" ω_{n-1} between the two equations involving m_1 and m_j gives:

$$x_n^j - m_j - (x_n - m_1)\frac{x_n^j - x_{n-1}^j}{x_n - x_{n-1}} = \sum_{i=1}^{n-2} \omega_i f(x_i)$$

$$m_j = x_n^j - (x_n - m_1)\frac{x_n^j - x_{n-1}^j}{x_n - x_{n-1}} + \sum_{i=1}^{n-2}\left(-f(x_i)\right)\omega_i$$

This clearly implies that

$$x_n^j - (x_n - m_1)\left(\frac{x_n^j - x_{n-1}^j}{x_n - x_{n-1}}\right) \le m_j.$$

It follows, as before, that for any fixed m_1, m_j is maximized by assigning as much weight as possible to x_1, i.e. by making ω_1 as big as possible. And again, for fixed weighted mean m_1, the minimum x_1 gets maximum weight when it is required to offset all by itself the maximum x_n. Recall that:

$$m_1 = \hat{\omega}_1 x_1 + (1 - \hat{\omega}_1)x_n \quad \Rightarrow \hat{\omega}_1 = \frac{x_n - m_1}{x_n - x_1}$$

and from this, we see that:

$$
\begin{aligned}
m_j &\le x_n^j - (x_n - m_1)\frac{x_n^j - x_{n-1}^j}{x_n - x_{n-1}} - \hat{\omega}_1 f(x_1) \\
&= x_n^j - (x_n - m_1)\frac{x_n^j - x_{n-1}^j}{x_n - x_{n-1}} - \left(\frac{x_n - m_1}{x_n - x_1}\right)\left(x_n^j - x_1^j - (x_n - x_1)\frac{x_n^j - x_{n-1}^j}{x_n - x_{n-1}}\right) \\
&= x_n^j - \left(\frac{x_n - m_1}{x_n - x_1}\right)\left(x_n^j - x_1^j\right) = x_n^j - (x_n - m_1)\left(\frac{x_n^j - x_1^j}{x_n - x_1}\right)
\end{aligned}
$$

We have shown that

$$
M_k(X) \subseteq \left\{ (1, m_1, \ldots, m_k) \;\middle|\; \begin{array}{c} x_1 \le m_1 \le x_n \\ x_n^j - (x_n - m_1)\left(\dfrac{x_n^j - x_{n-1}^j}{x_n - x_{n-1}}\right) \le m_j \le x_n^j - (x_n - m_1)\left(\dfrac{x_n^j - x_1^j}{x_n - x_1}\right) \; 2 \le j \le k \end{array} \right\}
$$

for $3 \le k \le n$, as required.

The point of this discussion, as regards using weights on a set X to fit pre-assigned moments, is that the number of elements of the set X limits the number of moments and the minimum and maximum values of the set X determines the allowable range of the moments. In particular, it may be advisable to arrange for X to encompass outliers, even at the expense of X being representative of claims experience, especially since it will be reweighted anyway and by design such outliers do not "adversely" impact the mean.

Section III: Finding the Weight

On the other hand, now suppose that (X, ω) was built to be representative of the kind of claims we are investigating and so we want to stay as "close as possible" to weight ω, in some sense.

Define the subset

$$\mathsf{P} = \left\{(\omega_1, \omega_2, \cdots, \omega_n) \mid 0 \le \omega_i, \sum_{i=1}^{n} \omega_i = 1\right\} \subset \mathsf{R}^n$$

that corresponds to probability weight functions. Note that this subset is closed, convex, and compact. Consider the $(k+1)\times n$ matrix (of maximum rank):

$$V_k = V_k(X) = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_n \\ \vdots & \vdots & \vdots & \vdots \\ x_1{}^k & x_2{}^k & \cdots & x_n{}^k \end{bmatrix}$$

Suppose we are given a vector $\hat{m} = (1, \hat{m}_1, \ldots \hat{m}_k)$ presumably derived from empirical data, and we are assured (or we refer to a characterization of covered moment vectors, as above, and augment the range of X if necessary) that the solution set

$$\mathsf{S} = \mathsf{P} \cap \left\{\upsilon \in \mathsf{R}^n \mid V_k \upsilon^T = \hat{m}^T\right\}$$

is not empty—in fact it is convex and compact. Since the norm function is continuous, it then follows that there is some $\upsilon_0 \in \mathsf{S}$ such that

$$\|\omega - \upsilon_0\| = Min \left\{\|\omega - \upsilon\| \mid \upsilon \in \mathsf{S}\right\}$$

making $\upsilon_0 \in \mathsf{S}$ in some sense an optimal reweighting of X, inasmuch as it fits the required moments while staying as close as possible to the original weight ω. We have verified that there exists a well-defined "best" solution, not necessarily unique, to the task of reweighting (X, ω) to fit a given set of moments.

Even though the set S is convex and compact and fairly well described, in general it is no picnic finding $\upsilon_0 \in \mathsf{S}$ that minimizes the distance to a point. We conclude this paper with two simpler approaches that, while lacking in theoretical appeal, are simple to implement.

Approach 1: Suppose, as above, it remains a priority to use a weight as near as practical with the original weight ω but we are only concerned with fitting the weighted mean to a given value $\hat{m}$, which we assume satisfies $x_1 < \hat{m} < x_n$. Consider the piecewise linear function:

$$f(\lambda,t) = Max(0, 2\lambda t + 1 - \lambda) \quad \lambda \in \mathrm{R}, t \in [0,1]$$

Notice the following limits:

$$\lim_{\lambda \to +\infty} f(\lambda,t) = \begin{cases} 0 & t \in [0,1) \\ +\infty & t = 1 \end{cases}$$

$$\lim_{\lambda \to -\infty} f(\lambda,t) = \begin{cases} 0 & t \in (0,1] \\ +\infty & t = 0 \end{cases}.$$

Consider the 1-parameter family of weights:

$$\omega_\lambda(x_i) = f\left(\lambda, \frac{x_i - x_1}{x_n - x_1}\right)\omega(x_i).$$

Note that $\omega_0 = \omega$. Define $g(\lambda) = \mu_1(X, \omega_\lambda)$. The reader can verify that g is a continuous, increasing function of λ with:

$$\lim_{\lambda \to -\infty} g(\lambda) = x_1 \quad \lim_{\lambda \to +\infty} g(\lambda) = x_n$$

It follows that there is a unique number λ_0 with $g(\lambda_0) = \hat{m}$. We remark that λ_0 can be readily found in practice with the use of a binary search algorithm. The weight $\upsilon = \frac{\omega_{\lambda_0} |X|_\omega}{|X|_{\omega_{\lambda_0}}}$ on X has the same total weight as the original weight ω with $\mu_1(X,\upsilon) = \mu_1(X, \omega_{\lambda_0}) = g(\lambda_0) = \hat{m}$.

Approach 2: Suppose, we are concerned with fitting both the weighted mean and variance, but it is not a priority to use a weight near the original weight ω. Suppose we are given a target mean $\hat{m}$ and variance $\hat{s}^2$. This approach exploits the Beta density, and we use the notation of [1]. We let $x_0 = 0 \le x_1$ and $L = x_n$. As in [1], the probability density function $g(z)$ of the two-parameter Beta density of mean $\hat{m}$ and variance $\hat{s}^2$ on the interval $(0, L)$ can be determined as:

$$c = \frac{\hat{s}}{\hat{m}} \qquad \alpha = \frac{L - \hat{m} - c^2\hat{m}}{Lc^2} > 0 \qquad \beta = \left(\frac{L - \hat{m}}{\hat{m}}\right)\alpha > 0$$

$$g(z) = g(\alpha, \beta; z) = \frac{z^{\alpha-1}(L-z)^{\beta-1}}{\mathrm{B}(\alpha,\beta)L^{\alpha+\beta+1}} \quad z \in (0, L).$$

Then define:

$$\upsilon(x_j) = \int_{x_{j-1}}^{x_j} g(z)dz \quad 1 \le j \le n.$$

We have:

$$|X|_\upsilon = \sum_{j=1}^{n} \upsilon(x_j) = \sum_{j=1}^{n} \int_{x_{j-1}}^{x_j} g(z)dz = \int_0^L g(z)dz = 1$$

and so υ is a probability weight on X. We also have:

$$\mu_k = \mu_k(X, \upsilon) = \frac{\sum_{x \in X} \upsilon(x) x^k}{|X|_\upsilon} = \sum_{j=1}^{n} x_j{}^k \int_{x_{j-1}}^{x_j} g(z)dz = \sum_{j=1}^{n} \int_{x_{j-1}}^{x_j} x_j{}^k g(z)dz$$

$$> \sum_{j=1}^{n} \int_{x_{j-1}}^{x_j} z^k g(z)dz = \int_0^L z^k g(z)dz$$

$$= \begin{cases} \hat{m} & k = 1 \\ \hat{m}^2 + \hat{s}^2 & k = 2 \end{cases}$$

Which indicates that while the weighted moments are greater, in most cases they should reasonably well approximate their target.

References:

1. Corro, Dan, *Fitting Beta Densities to Loss Data*, CAS Forum, Summer 2002.