An Adaptation of the Welford Method in order to Compute Semi-Variances

as adapted by Jean-Victor Côté

on January 8, 2012

Let x_k be real random variables, for $k \in \mathbb{N}$.

Define $T_{i,,j} = \sum_{k=i}^{j} x_k$, the sum of x_k, for $k \in \{i, ..., j\}$.

Define $M_{i,j} = \frac{1}{(j-i+1)} T_{i,j}$, the mean of x_k, for $k \in \{i, ..., j\}$.

For any $k \in \{2, ..., j\}$, it is clear that $x_k = kM_{1,k} - (k-1) M_{1,k-1}$, so that $(x_k - M_{1,k-1}) = k (M_{1,k} - M_{1,k-1})$, yielding the updating formula for means

$$M_{1,k} = M_{1,k-1} + \frac{1}{k} (x_k - M_{1,k-1})$$

by switching $kM_{1,k-1}$ from the right-hand side to the left-hand side of the original equation and dividing by k. We start with $M_{1,1} = x_1$and use the updating formula from there on.

Define $S_{i,j} = \sum_{k=i}^{j} (x_k - M_{i,j})^2$, the sum of squares of x_k, for $k \in \{i, ..., j\}$.

Let us find the updating formula for deviations $(x_k - M_{i,j})$.

For $k = j$, adding jx_j on both sides of $x_j = jM_{1,j} - (j-1) M_{1,j-1}$ and switching x_j and $jM_{1,j}$, we have that $jx_j - jM_{1,j} = (j-1) x_j - (j-1) M_{1,j-1}$, so that $(x_j - M_{1,j}) = (x_j - M_{1,j-1}) \left(\frac{j-1}{j}\right)$.

For $k < j$, in order to compute $(x_k - M_{1,j})$, we have to use the updating formula for means at j to find :

$$M_{1,j} = M_{1,j-1} + \frac{1}{j} (x_j - M_{1,j-1})$$

This yields

$$(x_k - M_{1,j}) = (x_k - M_{1,j-1}) - \frac{1}{j} (x_j - M_{1,j-1})$$

a formula in which we already have computed $M_{1,j-1}$ and $(x_k - M_{1,j-1})$, and in which $\frac{1}{j} (x_j - M_{1,j-1})$ has the same value across all $k < j$. Note that nowhere in this updating formula for deviations do we need the last updated mean.

If we only had to compute the variance, we could simplify the resulting summation formula as in [3]. Since we want to compute the semi-variance, we cannot, but we can update the deviations that we have already computed and add the last one to the list. All that remains to be done is to select the resulting deviations which are negative and divide the total of their squares by j , the number of observations in the sample so far, in order to obtain the last updated semi-variance. From one update to the next, we keep the last j, the last updated mean $M_{1,j}$ and the last updated list of deviations $\{x_k - M_{1,j}\}_{k=1}^{j}$.

References

[1] Chan, Tony F.; Golub, Gene H.; LeVeque, Randall k. (1983). Algorithms for Computing the Sample Variance: Analysis and Recommendations. The American Statistician 37, 242-247.

[2] Ling, Robert F. (1974). Comparison of Several Algorithms for Computing Sample Means and Variances. Journal of the American Statistical Association, Vol. 69, No. 348, 859-866.

[3] Welford, B.P. (1962). Note on a Method for Calculating Corrected Sums of Squares and Products. Technometrics, Vol.4, No.3, 419-420