Scientific Applications
F.N. Fritsch Editor

Updating Mean and Variance Estimates: An Improved Method

D.H.D. West
University of Dublin

A method of improved efficiency is given for updating the mean and variance of weighted sampled data when an additional data value is included in the set. Evidence is presented that the method is stable and at least as accurate as the best existing updating method.

Key Words and Phrases: mean, standard deviation, variance, updating estimates, removing data
CR Categories: 5.5, 5.19

1. Introduction

Given n data values $X_1, X_2, \ldots, X_n$, and corresponding weights $W_1, W_2, \ldots, W_n$, the weighted mean $\bar{X}$ and variance S^2 are defined by

$$\bar{X} = \left(\sum_{i=1}^{n} W_i X_i\right) \Big/ \sum_{i=1}^{n} W_i$$

and

$$S^2 = \left(\sum_{i=1}^{n} W_i (X_i - \bar{X})^2\right) \Big/ \left(\frac{n-1}{n} \sum_{i=1}^{n} W_i\right)$$

It is frequently convenient to calculate these quantities by a method that, unlike direct implementation of the above definitions, requires only one pass through the set of data pairs (X_i, W_i). Such a method is also useful where it is necessary to update (rather than recalculate *ab initio*) the mean and variance after the data set has been extended to include another data pair. A stable updating algorithm was given by Hanson [4]. An algorithm requiring considerably fewer multiplicative operations than Hanson's is given (WV2 in Table I), together with the results of numerical tests which indicate that this algorithm has numerical accuracy at least equal to that of Hanson's method. In a companion paper [1], Chan and Lewis give error bounds and the results of numerical tests for the unweighted versions of the present algorithm and several others; more extensive results, including theoretical and numerical comparisons between the unweighted forms of Hanson's algorithm and the present one, are to be found in their technical report [2].

Author's current address: D.H.D. West, The Harrison M. Randall Laboratory of Physics, The University of Michigan, Ann Arbor, MI 48109.

2. Algorithms

The algorithms which were numerically compared are given in Table I both as mathematical formulae and as informal program schemata. In the interests of simplicity of presentation, the updating algorithms (WV2-WV4) are given in a form in which the variance itself is calculated only at the completion of processing, and two practical considerations associated with the use of negative weights have been omitted: precautions against division by zero, and keeping a tally of the effective number of data pairs, since a negative weight corresponds to removal of a pair from the set. The appropriate changes are easily made if required.

Algorithm WV1 is a direct implementation of the definition of $\bar{X}$ and S^2; WV2 is proposed here as a replacement for WV3, which is Hanson's algorithm, and WV4 is the generalization to weighted data of the algorithm often given in statistics textbooks. WV2 may be obtained from WV3 by using for the calculation of the mean a different (and more accurate) method involving subexpressions which also appear in the variance updating formula, and which hence need not be recalculated. Similar methods can be applied to weighted linear regression [7]. The unweighted form of WV2 was also derived by Welford [6].

3. Numerical Tests

Each algorithm in Table I was run on artificial data sets of various sizes in which X_i and W_i were chosen independently from a normal distribution of mean unity and standard deviation σ. Data sets were generated having different values of sample size n and population standard deviation. The results, which are in each case averages over 20 different data sets with the same values of $\hat{X}$, S, and n, are shown in Tables II, III, and IV; κ is the condition number defined by Lewis and Chan [1] for the corresponding problem with unit positive weights. Negative values for the number of correct leading digits have been replaced by zero.

Removal of a data value from the set is sometimes accomplished by including it again but with the negative of its former weight. This process is unstable in finite

Table I. Algorithms for Weighted Variance S^2.

Algorithm	Mathematical	Computational
WV1 Two-Pass Algorithm	$\bar{X} = \frac{\sum_{i=1}^{n} W_i X_i}{\sum_{i=1}^{n} W_i}$ $S^2 = \frac{\sum_{i=1}^{n} W_i (X_i - \bar{X})^2}{\frac{n-1}{n} \sum_{i=1}^{n} W_i}$	SUMW = 0 SUMWX = 0 For i = 1, 2, . . . , n do { SUMW = SUMW + W_i { SUMWX = SUMWX + W_i * X_i XBAR = SUMWX/SUMW $S2$ = 0 For i = 1, 2, . . . , n do $S2$ = $S2$ + W_i * (X_i − XBAR) **2 $S2$ = $S2$ * n/((n − 1) * SUMW)
WV2 Proposed Algorithm	$M_1 = X_1$ $M_K = M_{K-1} + \frac{W_K}{\sum_{i=1}^{K} W_i} (X_K - M_{K-1})$ (K = 2, . . . , n) $\bar{X} = M_n$ $T_1 = 0$ $T_K = T_{K-1} + \frac{W_K}{\sum_{i=1}^{K} W_i} (X_K - M_{K-1})(X_K - M_{K-1}) \sum_{i=1}^{K-1} W_i$ (K = 2, . . . , n) $S^2 = \frac{T_n}{\frac{n-1}{n} \sum_{i=1}^{n} W_i}$	SUMW = W_1 M = X_1 T = 0 For i = 2, 3, . . . , n do { Q = X_i − M { TEMP = SUM + W_i { R = Q * W_i/TEMP { M = M + R { T = T + R * SUMW * Q { SUMW = TEMP XBAR = M $S2$ = T * n/((n− 1) * SUMW)
WV3 Hanson's Algorithm	$M_1 = X_1$ $M_K = \frac{M_{K-1} \sum_{i=1}^{K-1} W_i + W_K * X_K}{\sum_{i=1}^{K} W_i}$ (K = 2, . . . , n) $\bar{X} = M_n$ $T_1 = 0$ $T_K = T_{K-1} + \frac{W_K \sum_{i=1}^{K-1} W_i}{\sum_{i=1}^{K} W_i} (M_{K-1} - X_K)^2$ (K = 2, . . . , n) $S^2 = \frac{T_n}{\frac{n-1}{n} \sum_{i=1}^{n} W_i}$	SUMW = W_1 M = X_1 T = 0 For i = 2, 3, . . . , n do { TEMP = W_i * SUMW * (X_i − M) ** 2 { M = M * SUMW + W_i * X_i { SUMW = SUMW + W_i { M = M/SUMW { T = T + TEMP/SUMW XBAR = M $S2$ = T * n/((n − 1) * SUMW)
WV4 Textbook Algorithm	$\bar{X} = \frac{\sum_{i=1}^{n} W_i X_i}{\sum_{i=1}^{n} W_i}$ $S^2 = \frac{\sum_{i=1}^{n} W_i X_i^2 - \frac{\left(\sum_{i=1}^{n} W_i X_i\right)^2}{\sum_{i=1}^{n} W_i}}{\frac{n-1}{n} \sum_{i=1}^{n} W_i}$	SUMW = 0 SUMWX = 0 SUM$WX2$ = 0 For i = 1, 2, . . . , n do { SUMW = SUMW + W_i { TEMP = W_i * X_i { SUMWX = SUMWX + TEMP { SUM$WX2$ = SUM$WX2$ + TEMP * X_i XBAR = SUMWX/SUMW $S2$ = (SUM$WX2$ − SUMW * XBAR ** 2) * n/ ((n − 1) * SUMW)

precision arithmetic, but it has considerable practical utility, and tests were therefore run to compare the extent of the instability for different algorithms. Data sets were generated as before, except that for a sample of size n, data pairs $\frac{n}{2}$ + 2 through n were used to remove data pairs 2 through $\frac{n}{2}$ respectively. This is a very severe test for any algorithm. The results are shown in Tables V, VI, and VII.

Table II. Number of correct leading digits in S^2 for $n = 10$ when calculated by each algorithm.

σ	WV1	WV2	WV3	WV4	Condition number κ
1.0	8.0	7.9	8.0	7.6	1.6
10^{-1}	8.1	7.8	7.3	5.8	1.2×10^{1}
10^{-2}	8.1	6.6	6.5	3.6	1.2×10^{2}
10^{-3}	8.1	5.6	5.5	1.8	1.1×10^{3}
10^{-4}	7.8	4.7	4.3	0.0	1.2×10^{4}

Table III. Number of correct leading digits in S^2 for $n = 100$ when calculated by each algorithm.

σ	WV1	WV2	WV3	WV4	Condition number κ
1.0	7.6	7.7	7.7	7.4	1.4
10^{-1}	7.6	7.5	7.4	5.4	1.0×10^{1}
10^{-2}	7.7	6.6	6.5	3.5	1.0×10^{2}
10^{-3}	7.8	5.8	5.5	1.4	1.0×10^{3}
10^{-4}	7.1	4.8	4.6	0.0	1.0×10^{4}

Table IV. Number of correct leading digits in S^2 for $n = 1000$ when calculated by each algorithm.

σ	WV1	WV2	WV3	WV4	Condition number κ
1.0	7.1	7.2	7.1	6.9	1.4
10^{-1}	7.3	7.1	7.0	5.0	1.0×10^{1}
10^{-2}	7.2	6.7	6.5	3.1	1.0×10^{2}
10^{-3}	7.3	5.7	5.5	0.9	1.0×10^{3}
10^{-4}	6.2	4.6	4.5	0.0	1.0×10^{4}

Table V. Number of correct leading digits in S^2 for $n = 10$, with data removal, when calculated by each algorithm.

σ	WV1	WV2	WV3	WV4	Condition number κ
1.0	6.1	5.9	5.8	5.8	1.8
10^{-1}	5.8	4.9	4.7	3.5	1.2×10^{1}
10^{-2}	6.5	4.8	4.4	2.2	1.2×10^{2}
10^{-3}	5.1	1.9	1.7	0.0	1.1×10^{3}

Table VI. Number of correct leading digits in S^2 for $n = 100$, with data removal, when calculated by each algorithm.

σ	WV1	WV2	WV3	WV4	Condition number κ
1.0	2.7	3.3	2.3	1.9	1.4
10^{-1}	4.4	3.3	3.2	1.9	1.0×10^{1}
10^{-2}	4.7	2.6	2.6	0.6	1.0×10^{2}
10^{-3}	2.5	0.1	0.0	0.0	1.0×10^{4}

Table VII. Number of correct leading digits in S^2 for $n = 1000$, with data removal, when calculated by each algorithm.

σ	WV1	WV2	WV3	WV4	Condition number κ
1.0	3.2	3.1	2.9	2.6	1.4
10^{-1}	3.8	3.0	2.6	1.2	1.0×10^{1}
10^{-2}	2.8	0.1	0.1	0.0	1.0×10^{2}

Results for the accuracy in computing $\bar{X}$ are not given here since the differences between the algorithms were small in this case (less than 0.5 digit), and on difficult problems all algorithms gave better accuracy in $\bar{X}$ than any algorithm gave for S. The results in Tables II through VII were obtained on a DEC System 10 (27-bit mantissa). The same tests, with the same random numbers, were run on an Amdahl 470V/6 (6-hex-digit mantissa), with qualitatively similar results. The overall precision was about 2.5D lower, and the differences between algorithms were slightly greater on the Amdahl. The DEC-10 results are reproduced here because they cover a greater range of κ before all significance is lost.

4. Discussion and Conclusions

For the usual case of positive weights, the dependence of the error on n is weak for all the algorithms (both computers employed use guarded arithmetic—see [1, 2] for the expected behavior on machines which use unguarded arithmetic), and the dependence on κ is constant for WV1, linear for WV2 and WV3, and quadratic for WV4. These results parallel those for the corresponding unweighted algorithms [1, 2]. For negative weights, the results are more variable, but the relative behavior of the algorithms is similar to that for positive weights.

Execution time is usually dominated by the number of multiplications and divisions, and in this case the efficiency ranking of the algorithms is clear: WV4 leads with only 2 multiplications per data pair, followed by WV1 which has 3, WV2 with 4, and WV3 with 7. The numerical tests, however, show WV4 to be seriously unstable. The next most efficient algorithm, WV1, is highly stable and should always be used when two passes over the data are acceptable. When updating is required, the choice is therefore between WV3 (Hanson's algorithm) and WV2 (the new algorithm proposed here).

The forward error analyses of WV3 and WV2 are virtually identical, although a formal backward error analysis (i.e. an expression for error bounds in terms of equivalent perturbations in the data) exists for WV3 but apparently not for WV2 [5]. The numerical tests show that WV2 is slightly more accurate than WV3, and

that accordingly the new algorithm WV2 is to be preferred to WV3 when an updating method is required, because of its considerably greater efficiency.

Received October 1975; revised March 1979

References

1. Chan, T.F.C., and Lewis, J.G. Computing standard deviations: Accuracy. *Comm. ACM 22,* 9 (Sept. 1979), 526-531.
2. Chan, T.F.C., and Lewis, J.G. Rounding error analysis of algorithms for computing means and standard deviations. Tech. Rep. No. 289, Dept. of Mathematical Sciences, The Johns Hopkins U., Baltimore, Md., April 1978.
3. Cotton, I.W. Remark on stably updating mean and standard deviation of data. *Comm. ACM 18,* 8 (Aug. 1975), 458.
4. Hanson, R.J. Stably updating mean and standard deviation of data. *Comm. ACM 18,* 1 (Jan. 1975), 57-58.
5. Lewis, J.G. Private communication.
6. Welford, B.P. Note on a method for calculating corrected sums of squares and products. *Technometrics 4* (Aug. 1962), 419-420.
7. West, D.H.D. Incremental least squares and the approximate separation of exponentials. *Nuclear Instruments and Methods 136* (1976), 137-143.

Professional Activities Calendar of Events

ACM's calendar policy is to list open computer science meetings that are held on a not-for-profit basis. Not included in the calendar are educational seminars, institutes, and courses. Submittals should be substantiated with name of the sponsoring organization, fee schedule, and chairman's name and full address.

One telephone number contact for those interested in attending a meeting will be given when a number is specified for this purpose.

All requests for ACM sponsorship or cooperation should be addressed to Chairman, Conferences and Symposia Committee, Seymour J. Wolfson, 643 MacKenzie Hall, Wayne State University, Detroit, MI 48202, with a copy to Louis Fiora, Conference Coordinator, ACM Headquarters, 1133 Avenue of the Americas, New York, NY 10036; 212 265-6300. For European events, a copy of the request should also be sent to the European Representative. Technical Meeting Request Forms for this purpose can be obtained from ACM Headquarters or from the European Regional Representative. Lead time should include 2 months (3 months if for Europe) for processing of the request, plus the necessary months (minimum 3) for any publicity to appear in *Communications.*

■ This symbol indicates that the Conferences and Symposia Committee has given its approval for ACM sponsorship or cooperation.

In this issue the calendar is given in its entirety. New Listings are shown first; they will appear next month as Previous Listings.

NEW LISTINGS

13-16 January 1980
Optical Character Recognition Users Association Winter Conference, Orlando, Fla. Sponsor: OCR Users Association. Contact: OCR Users Association, 10 Banta Place, Hackensack, NJ 07601; 201 343-4935.

18-20 March 1980
Electric Power Problems: The Mathematical Challenge, Seattle, Wash. Sponsor: SIAM. Contact: Albert M. Erisman, Boeing Computer Services Co., 565 Andover Park West, M/S 9C-01, Tukwila, WA 98188.

8-11 April 1980
Fifth European Meeting on Cybernetics and Systems Research, Vienna, Austria. Sponsor: Austrian Society for Cybernetic Studies. Contact: Conference Secretariat, Schottengasse 3, A-1010 Vienna, Austria.

13-16 April 1980
AEDS Annual Convention, St. Louis, Mo. Sponsor: AEDS. Contact: Ralph Lee, University of Missouri-Rolla, Rolla, MO 65559.

April 28-30, 1980
■ **Twelfth Annual ACM Symposium on Theory of Computing,** Los Angeles, Cal. Sponsors: ACM SIGACT and University of Southern California. Symp. chm: Richard J. Lipton, College of Engineering, University of California, Berkeley, CA 94720.

5-7 May 1980
TIMS/ORSA National Meeting, Washington, D.C. Sponsors: TIMS/ORSA. Contact: Donald Gross, School of Engineering, George Washington University, Washington, DC 20052.

11-14 May 1980
ASM 1980 Annual Conference, New Orleans, La. Sponsor: Association for Systems Management. Contact: R.B. McCaffrey, ASM, 24587 Bagley Road, Cleveland, OH 44138.

17-19 June 1980
Computerized Energy Management Control Systems Conference, Tulsa, Okla. Sponsor: Oklahoma State University, Oklahoma Dept. of Energy. Contact: Wayne C. Turner or Phillip M. Wolfe, School of Industrial Engineering and Management, 322 Engineering North, OSU, Stillwater, OK 74074.

2-4 July 1980
Conference on Databases, University of Aberdeen, Aberdeen, Scotland. Sponsor: University of Aberdeen. Contact: S.M. Deen, Dept. of Computing Science, University of Aberdeen, Old Aberdeen AB9 2UB, Scotland, U.K.

14-18 July 1980
■ **SIGGRAPH 80, Seventh Annual Conference on Computer Graphics and Interactive Techniques,** Seattle, Wash. Sponsor: ACM SIGGRAPH. Contact: SIGGRAPH 80, Box 88203, Seattle, WA 98188; 206 453-0599.

1-5 September 1980
9th International Symposium on Mathematical Foundations of Computer Science, Rydzyna, Poland. Sponsors: Polish Academy of Sciences Institute of Computer Science. Symp. chm: Jan Maluszynski, Institute of Computer Science, Polish Academy of Sciences, Box 22, 00-901 Warsaw PKiN, Poland.

9-11 November 1981
■ **ACM Annual Conference,** Los Angeles, Calif. Sponsor: ACM. Conf. chm: A.C. (Toni) Shetler, Secretary Development Staff, Xerox Corp. A3-49, 701 South Aviation Blvd., El Segundo, CA 90245; 213 679-4511 x1968.

PREVIOUS LISTINGS

16-18 September 1979
Nonbibliographic Databases, Boston, Mass. Sponsor: Association of Information and Dissemination Centers. Contact: ASIDIC, Box 8105, Athens, GA 30603.

16-22 September 1979
Lambda-Calculus Conference, University College, Swansea, U.K. Contact: R. Hindley, Lambda Conference, Pure Mathematics Dept., University College, Swansea SA2 8PP, U.K.

17-19 September 1979
■ **4th International Conference on Software Engineering,** Munich, Fed. Republic of Germany. Sponsors: ACM SIGSOFT, IEEE-CS, ERO, Gesellschaft für Informatik. Conf. chm: Frederick L. Bauer, Institut für Informatik, Technische Universitat Munchen, Munich, W. Germany.

17-20 September 1979
Design Research Society Conference, University of Bristol, U.K. Sponsor: DRS. Conf. Sec: R. Gill, 91 Woodland Road, Bristol BS8 1US, U.K.

20-21 September 1979
Sorrento Workshop for International Standardization of Simulation Languages (before IMACS Congress Sept. 24-28), Sorrento, Italy. Contact: Tuncer I. Ören, Computer Science Dept., University of Ottawa, Ottawa, Ont. K1N 6N5, Canada.

20-26 September 1979
3rd World Telecommunication Exhibition, Geneva, Switzerland. Sponsor: International Telecommunication Union. Contact: Secretariat TELECOM 79, Orgexpo, 18 quai Ernst-Ansermet, Case postale 65, 1211 Geneva 4, Switzerland.

21-22 September 1979
Conference on Database Technology, Bad Nauheim, Kurhaus, Germany. Sponsor: ACM German Chapter. Contact: J. Niedereichholz, University of Frankfurt, Institut für Wirtschaftsinformatik, Mertonstrasse 17-25, D 6000 Frankfurt/Main, Germany.

24-28 September 1979
5th IFAC Symposium on Identification and System Parameter Estimation, Darmstadt, Fed. Republic Germany. Sponsor: IFAC. Contact: IFAC 1979, c/o VDI/VDE—Gesellschaft Mess- und Regelungstechnik, Postfach 1139, D-4000 Düsseldorf 1, Federal Republic Germany.

25-28 September 1979
Euro IFIP 79, London, U.K. Sponsor: IFIP. Contact: Euro IFIP 79, IFIP Foundation, Paulus Potterstraat 40, 1071 DB Amsterdam, The Netherlands.

26-28 September 1979
International Symposia: Computers and Education, Applications and Software Engineering, Montreal, Canada. Sponsor: IASTED in cooperation with ISMM, Acta Press. Contact: The Secretary, IASTED/ISMM Symposia, Box 2481, Anaheim, CA 92804.

26-28 September 1979
Computers in Cardiology, Geneva, Switzerland. Sponsors: American Heart Association, European Society of Cardiology, NIH, IEEE-CS. Contact: Computers in Cardiology, Centre de Cardiologie, Hôpital Cantonal, 1211-Geneva 4, Switzerland.

26-29 September 1979
Ninth International Symposium and Exhibition, Mini and Microcomputers, Montreal, Canada. Sponsor: International Society for Mini and Microcomputers. Contact: The Secretary, MIMI 79 Montreal, Box 2481, Anaheim, CA 92804.

27-28 September 1979
■ **2nd SIGIR Conference,** Dallas, Tex. Sponsor: ACM SIGIR. Conf. chm: Robert R. Korfhage, Dept. of Computer Science, Southern Methodist University, Dallas, TX 75275; 214 692-3082.

30 September-3 October 1979
■ **SIGUCC User Services Conference VII,** Los Angeles, Calif. Sponsor: ACM SIGUCC (University Computing Centers). Conf. chm: Jerome A. Smith, Acting Director, Computing Services, 269 Evans Hall, University of California, Berkeley, CA 94720; 415 642-0889.

1-3 October 1979
■ **Second Annual Symposium on Small Systems,** Dallas, Tex. Sponsor: ACM SIGSMALL. Conf. chm: Gerald Kane, CS and EE Dept., 214 Patterson Hall, Southern Methodist University, Dallas TX 75275; 214 692-3081.

1-4 October 1979
1st International Conference on Distributed Computing Systems, Huntsville, Ala. Sponsor: US Army Ballistic Missile Defense Advanced Technology Center in cooperation with IEEE-CS, IRIA, IPSJ. Gen. co-chm: Charles R. Vick, BMD-ATC, Data Processing Directorate, Box 1500, Huntsville, AL 35807; 205 895-4175.

1-6 October 1979
Informatica 79, Bled, Yugoslavia. Sponsor: Slovene Computer Society in cooperation with Jozef Stefan Institute, University of Ljubljana. Contact: Informatica 79, Institut Jozef Stefan, 61001 Ljubljana, pp. 199, Yugoslavia.

2-5 October 1979
ECOMA-7, Capacity Management: Techniques and Implementation in the 80s, Paris, France. Sponsor: European Computer Measurement Association. Contact: Scott N. Yasler, European Computer Measurement Association (Dept. CFP), Scheuchzerstrasse 5, CH-8006 Zurich, Switzerland.

3-5 October 1979
■ **Fifth International Conference on Very Large Data Bases,** Rio de Janeiro, Brazil. Sponsors: ACM SIGIR, SIGMOD, and SIGBDP, IEEE-CS, IFIPS, U.S. Conf. chm: Stanley Y.W. Su, Computer & Information Sciences Dept., University of Florida, Gainesville, FL 32611.

8-9 October 1979
■ **1979 International Symposium on Computer Hardware Description Languages and Their Applications,** Palo Alto, Calif. Sponsors: ACM SIGDA, IEEE-CS. Prog. chm: Donald L. Dietmeyer, Dept. of Electrical and Computer Engineering, University of Wisconsin, 1425 Johnson Drive, Madison, WI 53706; 608 262-3890.

10-12 October 1979
Seventeenth Annual Allerton Conference on Communication, Control, and Computing, Allerton House, near Monticello, Ill. Sponsor: University of Illinois at Urbana-Champaign. Conf. co-chm: J.B. Cruz Jr. and F.P. Preparata, Dept. of EE and Coordinated Science Laboratory, University of Illinois at Urbana-Champaign, Urbana, IL 61801.

10-12 October 1979
AICA Annual Congress, Bari, Italy. Sponsor: Assoziazione Italiana per il Calcolo Automatico. Congr. Sec: Maria Teresa Pazienza, Corso di Laurea in Scienze dell'Informazione-Istituto di Fisica, Via Amendola, 173-Bari, Italy.

14-17 October 1979
28th International Conference and Business Exposition, San Diego, Calif. Sponsor: DPMA. Contact: Carol Harte, DPMA, 505 Busse Highway, Park Ridge, IL 60068.

(Calendar continued on p. 537)