Comp. & Maths. with Appls. Vol. 9, No. 2, pp. 317-322, 1983
Printed in Great Britain.
0097-4943/83/020317-06$03.00/0

ERROR ANALYSIS OF AITKEN's Δ^2 PROCESS

M. PETER JURKAT
Department of Management Science, Stevens Institute of Technology, Castle Point, NJ 07030, U.S.A.

(*Received August* 1982)
Communicated by E. Y. Rodin

Abstract—A sequence to sequence transformation, called the Δ^2 process by its developer Aitken, and recently analyzed by Daniel Shanks, is successful in accelerating convergence in many convergent sequences and inducing convergence in some divergent ones. It is shown here that the Δ^2 process applied to sequences whose terms have Cauchy distributions results in sequences whose terms still have the Cauchy distribution and that repeated applications of the Δ^2 process to a sequence with terms having uniform distribution (simulating round-off error) and to sequences with terms having a normal distribution (simulating measurement error) yields, in both cases, sequences whose terms approach the Cauchy distribution. The result for the uniform distribution is proven, that for the normal distribution is referenced.

Daniel Shanks[2] has described a family of sequence-to-sequence transformation with the property that they could accelerate convergence in many convergent sequences and induce convergence in divergent ones. The basic transformation, also known as Aitken's[1] Δ^2 process, is defined by

$$e_1(a_n) = (a_{n+1}a_{n-1} - a_n^2)/(a_{n+1} + a_{n-1} - 2a_n) \tag{1}$$

for $n = 1, 2, \ldots$ where $\{a_n\}$, $n = 0, 1, \ldots$ is a sequence of numbers, possibly the partial sums of a series. The elements of Shank's family of transformation may be constructed from e_1, by iteration or the selection of subsequences from, say, $e_k(a_n) = e_1{}^k(a_n)$. These transformations work particularly well for the partial sums of series approximating geometric series. The sum of $1 + x + x^2 + \ldots$ may be calculated from e_1 applied to the first three terms; thus $e_1(a_2) = 1/(1-x)$.

The object here is to classify the distribution of $e_1(a_n)$ when the a_n are selected from various statistical distributions. The analysis is performed as if the general terms of the sequence are defined as

$$a_n = 0 + \alpha_n$$

where α_n is a random variable with some assumed distribution. This provides information on a sequence consisting of "pure error". The application of these results to non-zero sequences with error amounts to a term-by-term shift in the lowest order location parameter of the assumed distribution, often the mean.

The research was motivated by the recent increase in numerical simulation exercises, particularly those that iteratively solve the equations of a model of a system being driven to a steady state. The parameters of the system are often known only to within some error while the large number of calculations involved make round-off error of some concern. The former type of error is here modeled by normal distribution while the latter by the uniform distribution. Initially no assumption about the form of the α_n's is made.

To simplify notation, and without loss of generality, the analysis is performed on $y = e_1(a_2)$.

GENERAL FORMULA

In this section the a_i's will be considered as random variables with well defined but arbitrary distributions. Desired is $H(z) = P\{y < z\}$, where

$$y = (a_3a_1 - a_2^2)/(a_3 + a_1 - 2a_2).$$

Rewrite

$$y = (a_1 a_3 - a_2^2)/(a_1 + a_3 - 2a_2) = \frac{\begin{vmatrix} a_1 & a_2 - a_1 \\ a_2 & a_3 - a_2 \end{vmatrix}}{\begin{vmatrix} 1 & a_2 - a_1 \\ 1 & a_3 - a_2 \end{vmatrix}}.$$

This may be considered the formula for one of the solutions, by Cramer's rule, of

$$C_1 : y + x(a_2 - a_1) = a_1$$

$$C_2 : y + x(a_3 - a_2) = a_2$$

The two lines, C_1 and C_2, meet at some point, R, whose y-coordinate is $e_1(a_2)$. To determine $P\{y < z\}$, one may determine from given values of a_1 and a_2 what values of a_3 keep R below the horizontal $y = z$.
Figure 1 shows that:

—a_1 is the y-intercept of C_1
—a_2 is the y-intercept of C_2 and the y-value of C_1 at $x = -1$
—a_3 is the y-value of C_2 at $x = -1$

Thus, C_2 may be rotated about $(0, a_2)$ so that R stays below $y = z$ and a_3 can be read off C_2 at $x = -1$. Define $A = 2a_2 - a_1$ as the y-value at $x = -1$ of the line C_3 through $(0, a_2)$ parallel to C_1 and $B = a_2 + (a_2 - a_1)(a_2 - z)/(a_1 - z) = z + (a_2 - z)^2/(a_1 - z)$ as the y-value at $x = -1$ of a line through $R_1 = C_1 \cdot \{y = z\}$ and $(0, a_2)$. Let C_4 designate the line through R_1 and $(0, a_2)$. Then C_3

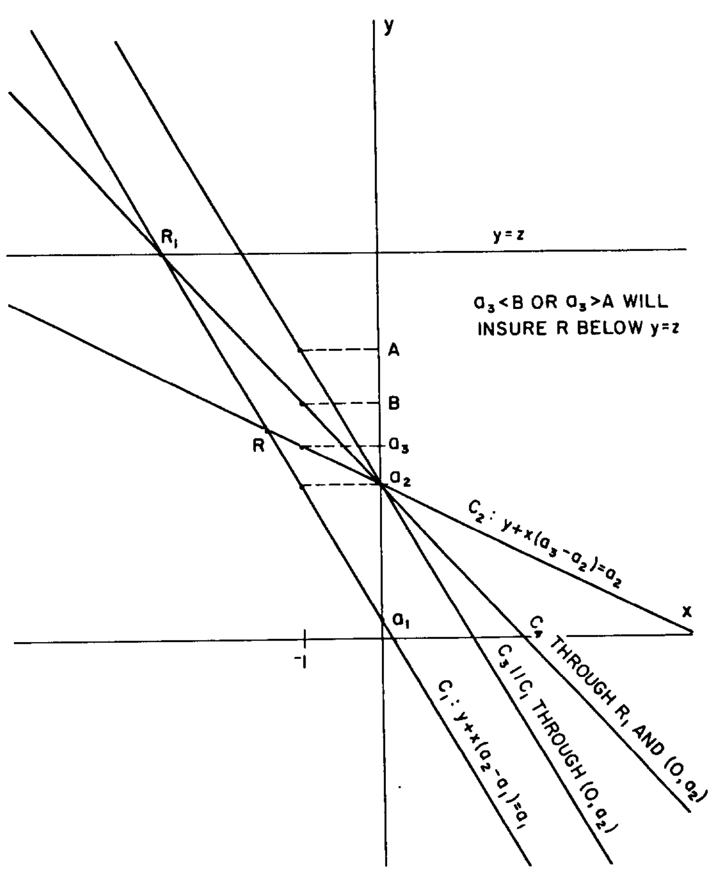

Fig. 1. Geometric determination of the a_3-integration limits for $z > a_2 > a_1$.

and C_4 are the boundaries for a set of lines through $(0, a_2)$ for which R is below $y = z$. Thus $a_3 = A$ and $a_3 = B$ are the boundaries for a_3 such that $y < z$ for a given a_1 and a_2.

Let $F(a)$ be the cumulative distribution function of a_1, a_2 and a_3. Several cases may be distinguished. In Fig. 1, the case $z > a_2 > a_1$ is shown. This case, and the case for $a_2 > z > a_1$ yield the same result, so

$$P\{y < z | a_2 > a_1, z > a_1\} = \int_{-\infty}^{z} \int_{a_1}^{\infty} \left[1 - \int_B^A dF(a_3)\right] dF(a_2) dF(a_1)$$

where the verticle line may be read as "for the case where".

Similar arguments, with slightly different geometric conditions, show that

$$P\{y < z | a_2 > a_1 > z\} = \int_z^{\infty} \int_{a_1}^{\infty} \int_A^B dF(a_3) dF(a_2) dF(a_1)$$

$$P\{y < z | a_1 > a_2, a_1 > z\} = \int_z^{\infty} \int_{-\infty}^{a_1} \int_A^B dF(a_3) dF(a_2) dF(a_1)$$

$$P\{y < z | z > a_1 > a_2\} = \int_{-\infty}^{z} \int_{-\infty}^{a_1} \left[1 - \int_B^A dF(a_3)\right] dF(a_2) dF(a_1).$$

The four cases are mutually exclusive, so the probability may be added and like integrals combined to yield the general formula for the cumulative distribution of y

$$H(z) = \int_{-\infty}^{z} \int_{-\infty}^{\infty} \left[1 - \int_B^A dF(a_3)\right] dF(a_2) dF(a_1) + \int_z^{\infty} \int_{-\infty}^{\infty} \int_A^B dF(a_3) dF(a_2) dF(a_1). \quad (2)$$

Notice that since a_1 and a_3 enter the formula for y symmetrically, their roles in $H(z)$ may be interchanged.

To see that H is a proper c.d.f., one notices that lim $(B - A) = 0$ as $z \to -\infty$ and lim $(A - B) = 0$ as $z \to \infty$. Then as $z \to -\infty$, the first term of H is zero due to the restriction of the range of a_1 and the second is zero due to the restriction of the range of a_3. Hence $H(-\infty) = 0$. As $z \to \infty$, the second term goes to zero and the first goes to

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} [1 - 0] dF(a_2) dF(a_1) = 1$$

since F is a c.d.f. Hence $H(\infty) = 1$. If F has a frequency function $f(a) = F'(a)$ then

$$h(z) = H'(z) = \int_{-\infty}^{\infty} f(a_1) \int_{-\infty}^{\infty} f(a_2) f(B)(a_1 - a_2)^2/(a_1 - z)^2 da_2 da_1$$

which is non-negative, since f is. So H is non-decreasing and hence a c.d.f.

ROUNDOFF ERROR

As is common, roundoff error is assumed distributed uniformly. Let, for $h > 0$,

$$f(a) = 1/(2h) \text{ for } -h \le a \le h$$
$$= 0 \text{ elsewhere.}$$

The problem in evaluating $H(z)$ is that the discontinuities of f at $-h$ and h force the limits of integration to be split into many segements for each of the a_1, a_2, a_3 integrations.

This integration was performed and is reported in near exhausting detail in Jurkat[3]. The results are given for z rewritten as $z = hy$. In dimensionless form, then, $G(y) = H(hy)$ is

$$G(y) = -y/2 - ([1 - y]^{3/2} - [-1 - y]^{3/2})^2/18 \qquad y < -1$$

$$G(y) = (y^2 + y + 1)/2 - ([1 - y]^3 - [1 + y]^3)/18 + \frac{1}{12}(1 + y)^3 \log\left(\frac{1 - y}{1 + y}\right) \qquad -1 \le y < 0$$

$$G(y) = 1 - (y^2 - y + 1)/2 - [(1+y)^3 - [1-y]^3]/18 + \frac{1}{12}(1-y)^3 \log\left(\frac{1+y}{1-y}\right) \qquad 0 \le y < 1$$

$$G(y) = 1 - y/2 + ([1+y]^{3/2} - [-1+y]^{3/2})/18 \qquad 1 \le y.$$

These formulas, although numerically verified, are not readily comprehended so approximations were sought.

Numerical calculation indicated that the tails of G contain more probability than any normal distribution, so an approximation with a Cauchy distribution was attempted. Let $c(y;\ \beta) = (\beta/\pi)/(1/(\beta^2 + y^2))$ be the Cauchy frequency function and write

$$G(y) = 1 - y/6 + y^3/9 - (-1 + y^2)^{3/2}/9 \text{ for } 1 \le y$$

then

$$g(y) = G'(y) = -1/6 + y^3/3 - y(-1 + y^2)^{1/2}/3$$

and

$$\lim_{y \to \pm\infty} \frac{g(y)}{c(y;\ \beta)} = \pi/24\beta.$$

This means that $g(y) \sim 24/(\pi^2 + (24y)^2)$. The approximation is valid to within 2% already at $y = 1$ since $G(1) = 0.9\bar{4}$ and the Cauchy cumulative distribution $C(1;\ 1/7) = 0.9548$ and $C(1;\ 1/8) = 0.9604$ $[1/8 < \pi/24 < 1/7]$.

A Monte Carlo "simulation" of the random process described by e_1 was made. One thousand triplets of pseudo-random numbers distributed uniformly in the interval $[-1, 1]$ were selected using the RAN routine built into the PDP-10 FORTRAN compiler. The formula e_1 was then applied to each triplet and tabulated. The results are indicated in the table. Also shown in the table are the results of successive application of e_1 to the results of the previous Monte Carlo simulation. The first quadrant values are plotted on Fig. 2 along with $C(y;\ 1)$. It may be

Table 1. Comparison of direct calculation and successive Monte Carlo simulation of e_1 applied to uniform variates

y	Direct Calculation of G (y)	Monte Carlo Simulations				
		e_1	e_2	e_3	e_4	e_5
-7	.0060	.0070	.0010	.0190	.0140	.0100
-6	.0070	.0090	.0130	.0200	.0150	.0120
-5	.0084	.0100	.0130	.0220	.0190	.0130
-4	.0105	.0130	.0150	.0260	.0220	.0270
-3	.0142	.0190	.0200	.0330	.0290	.0310
-2	.0218	.0230	.0320	.0450	.0440	.0430
-1	.0556	.0460	.0760	.0860	.0790	.0870
-.5	.2059	.2000	.1750	.1570	.1510	.1690
0	.5000	.5120	.4950	.4880	.4810	.4970
1	.9444	.9470	.9300	.9230	.9210	.9190
2	.9782	.9830	.9720	.9640	.9680	.9630
3	.9858	.9890	.9850	.9850	.9760	.9720
4	.9895	.9920	.9880	.9900	.9820	.9780
5	.9916	.9970	.9920	.9900	.9850	.9840
6	.9930	.9970	.9950	.9900	.9880	.9860
7	.9940	.9970	.9970	.9920	.9900	.9910

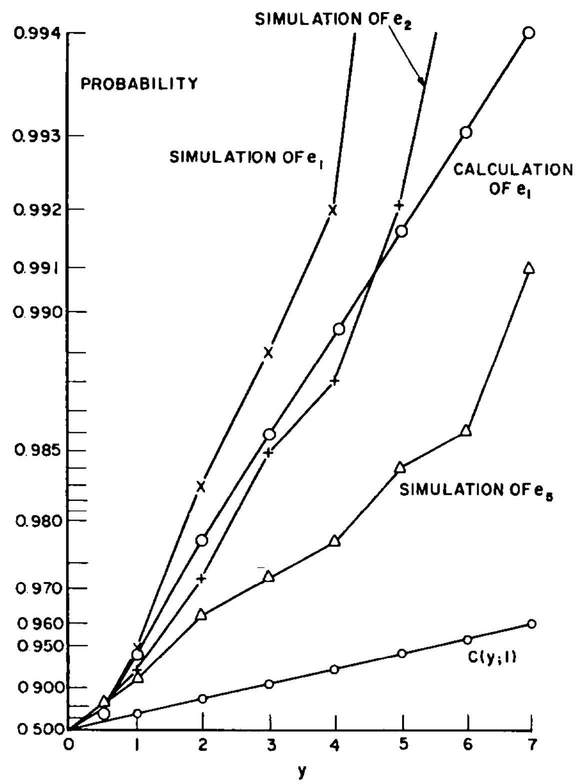

Fig. 2. Results of Monte Carlo simulation of e_1 applied to the uniform distribution.

noticed that the initial simulation of e_1 agrees well with the explicit calculation of $G(y)$ but successive iteration on e_1 deviates further from G in the direction of $C(y; 1)$.

The approximations and calculations indicated that e_1 transforms sequences whose elements are uniformly distributed into sequences whose elements are nearly Cauchy. Numerical calculations reported in Jurkat[3] showed a similar result for sequences with elements distributed normally. The question then arises as to what happens to sequences whose elements are themselves distributed Cauchy. The result is that the Cauchy distribution acts like a fixed point in the space of distribution under the transformation e_1. The confirmation of this is given in the next section.

CAUCHY DISTRIBUTION

Let each element of the sequence $\{a_i\}$ have the frequency function

$$f(a_i) = \frac{1}{\pi}\frac{1}{1+a_i^2} \quad i = 1, 2, 3.$$

Performing the a_3 integration explicitly, the frequency function of y is

$$hy) = \frac{1}{\pi^3}\int_{-\infty}^{\infty}\frac{1}{1+a_1^2}\int_{-\infty}^{\infty}\frac{1}{1+a_2^2}\frac{1}{1+\left(y+\frac{(a_2-y)^2}{a_1-y}\right)^2}\left(\frac{a_1-a_2}{a_1-y}\right)^2 \mathrm{d}a_2\,\mathrm{d}a_1$$

$$= \frac{1}{\pi^3}\int_{-\infty}^{\infty}\frac{1}{1+(a_1+y)^2}\int_{-\infty}^{\infty}\frac{1}{1+(a_2+y)^2}\frac{(a_1-a_2)^2}{(1+y^2)a_1^2+2ya_1a_2^2+a_2^4}\mathrm{d}a_2\mathrm{d}a_1$$

If the $(a_1-a_2)^2$ term in the numerator is expanded and

$$h(y, k) = \frac{1}{\pi^3(1+y^2)}\int_{-\infty}^{\infty}\frac{1}{1+(a_1+y)^2}\int_{-\infty}^{\infty}\frac{1}{1+(a_2+y)^2}\frac{k}{\left(a_1+\frac{ya_2^2}{1+y^2}\right)^2+\frac{a_2^4}{(1+y^2)^2}}\mathrm{d}a_2\mathrm{d}a_1$$

then

$$h(y) = h(y, a_1^2) - 2h(y, a_1 a_2) + h(y, a_2^2).$$

Each of the terms of this expression is evaluated separately and recombined.

The evaluation is done by residues. There are only a finite number of poles of the integrands in $h(y, k)$ so there is one with maximum modulus. Let S be a real number greater than the maximum modulus of the poles of the integrands of $h(y, k)$. For $R > S$ define the contour T by

$$T = T_1 + T_2 \text{ where } T_1 = [-R, R]$$
$$T_2 = \mathrm{Re}^{i\theta} \text{ for } \theta \in [0, \pi]$$

Since all the poles with positive imaginary part of the integrands of $h(y, k)$ are within T, increasing the value of $R > S$ does not change the value of the sum of the residues. Since all integrands are rational functions (with the denominator degree of 4 in a_1 and 6 in a_2 while the numerator degree is only 2 in both) the integrals over T_2 approach zero as $R \to \infty$.

Upon finding the residues of each of the terms of $h(y)$, it may be shown that

$$h(y) = \frac{1}{\pi}\frac{y^2+2}{4(y^2+1)} - \frac{1}{\pi}\frac{y^2}{2(y^2+1)} + \frac{1}{\pi}\frac{y^2+2}{4(y^2+1)} = \frac{1}{\pi}\frac{1}{y^2+1}$$

which is the frequency function of the Cauchy distribution that was assumed to be the distribution of the individual a_i in $h(y, k)$.

This shows that the Cauchy distribution is a fixed point under the transformation e_1.

REFERENCES

1. A. C. Aitken, On Bernoulli's numerical solution of algebraic equations. *Proc. Roy. Soc. Edinburgh* **46**, 289–305 (1926).
2. Daniel Shanks, Non-linear transformations of divergent and slowly convergent sequences. *J. Math. Phys.* **34**(1) (1955).
3. M. P. Jurkat, Error analysis in the determination of D. C. level by Shank's method. Ph.D. Dissertation, Stevens Inst. of Tech., 1972.