ON A DIFFERENCE EQUATION ARISING IN A LEARNING-THEORY MODEL

BY

BERNARD EPSTEIN*

ABSTRACT

An analysis is presented of the equation $f(x + a) - f(x) = e^{-x} \{f(x) - f(x-b)\}$. Here a and b denote arbitrary positive constants, and a solution is sought which satisfies the following conditions: $f(-\infty) = 0, f(+\infty) = 1$, $0 \leqq f(x) \leqq 1$. Existence and uniqueness of solution are established, and then an analytical form of the solution is obtained by use of bilateral Laplace transform.

1. This paper is devoted to a study of the difference equation

$$f(x+a) - f(x) = e^{-x}\{f(x) - f(x-b)\}, \tag{1}$$

which arose in connection with a certain model for a learning process. (See [1], [2].) Here a and b denote positive constants and $f(x)$ denotes, for each value of x, a probability of absorption, so that the inequalities $0 \leqq f(x) \leqq 1$ must be satisfied. The appropriate boundary conditions are given by $f(-\infty) = 0$, $(+\infty) = 1$. Since f does *not* denote a cumulative distribution function, monotonicity is not required. However, it will be shown that there exists a unique bounded function which satisfies (1) and the prescribed conditions at $\pm\infty$, and that this solution is monotone; thus, the inequalities $0 \leqq f(x) \leqq 1$ will certainly be satisfied.

It is immediately seen that (1) possesses at most one continuous solution satisfying the boundary conditions. For, if $g(x)$ were the difference of two distinct such solutions, it would satisfy (1) and vanish at $\pm\infty$, so that it would attain a positive maximum (or negative minimum) on a non-empty compact set S. Setting x in (1) (with f replaced by g) equal to the maximum of S, one immediately obtains the desired contradiction. A simple modification of this argument suffices to show that uniqueness also holds under the weaker hypothesis of boundedness.

Suppose now that $f(x)$ is a function of bounded variation which satisfies (1) and the boundary conditions. From the continuity of the factor e^{-x} it follows that the right-hand limit, $f(x+0)$, also satisfies (1), and the same is then true of the right-hand discontinuity,

Received, March 8, 1966.

* Research supported by the National Science Foundation, Grant GP-2558.

$$s(x) = f(x+0) - f(x). \tag{2}$$

If $f(x)$ possessed any right-hand discontinuities, $s(x)$ would assume non-zero values on a non-empty set $\{x_1, x_2, \cdots\}$, finite or denumerable, and the inequality

$$\Sigma \left| s(x_k) \right| < \infty \tag{3}$$

would hold. Assuming that $s(x)$ takes positive values, we choose for x the point at which it attains its maximum (or the largest of these, if there are several); we thus obtain $s(x+a) - s(x) < 0$, $s(x) - s(x-b) \geqq 0$, contradicting (1) (with $f(x)$ replaced by $s(x)$). If $s(x)$ is supposed to assume negative and zero values only, a similar argument is employed. Therefore, $s(x)$ must vanish identically; similarly, $f(x)$ possesses no left-hand discontinuities, and so it must be continuous.

We shall now obtain, by two different methods, a monotone and a continuous solution, respectively, of (1), both satisfying the boundary conditions. It will then follow from the above remarks that the two solutions coincide.

2. In this section we shall construct by iteration a monotone solution of (1). It will be convenient to re-write (1) in the form

$$f(x) = w(x)f(x+a) + (1 - w(x))f(x-b), \tag{4}$$

where $w(x) = e^x/(1+e^x)$. It is natural to attempt to construct a solution by iteration, beginning with a more or less arbitrary initial function $f_0(x)$:

$$f_{n+1}(x) = w(x)f_n(x+a) + (1 - w(x))f_n(x-b), \qquad n \geqq 0. \tag{5}$$

It is immediately evident that if $f_0(x)$ satisfies the boundary conditions, the same will be true of all the succeeding functions $f_1(x)$, $f_2(x), \cdots$. The monotonicity and the positivity of the factors $w(x)$ and $1 - w(x)$ are readily seen to imply that, if $f_0(x)$ is monotone, all the succeeding functions will also possess this property. Finally, the positive character of the factors $w(x)$ and $1 - w(x)$ evidently guarantees that if $f_0(x)$ is so chosen that $f_1(x) \geqq f_0(x)$ holds everywhere, then, more generally, $f_{n+1} \geqq f_n(x)$ will also hold everywhere; similarly if the inequalities are reversed.

Now suppose that we can find monotone functions, $f_0(x)$ and $g_0(x)$, both satisfying the boundary conditions, such that the inequalities $f_0(x) \leqq f_1(x)$, $g_0(x) \geqq g_1(x)$ hold everywhere. Then it is immediately evident that the sequences $f_0(x), f_1(x), \cdots$ and $g_0(x), g_1(x), \cdots$ converge pointwise to monotone functions, $f(x)$ and $g(x)$ respectively, which satisfy (1). If, furthermore, $f_0(x) \leqq g_0(x)$ holds everywhere, it then follows immediately from (5) (and the analogous equation for the g's) that $f_n(x) \leqq g_n(x)$ everywhere. Thus, $f(x) \leqq g_0(x)$, and so

$$0 \leqq f(-\infty) \leqq g_0(-\infty) = 0. \tag{6a}$$

Similarly,

$$1 = g_0(+\infty) \geqq f(+\infty) \geqq f_0(+\infty) = 1. \tag{6b}$$

Thus, $f(x)$ satisfies (1) and the prescribed boundary conditions; similarly for $g(x)$, and so, by the introductory remarks, $f(x) = g(x)$.

It remains to demonstrate a pair of functions $\{f_0(x), g_0(x)\}$ satisfying the conditions imposed in the preceding paragraph. Simple calculations show that the following choices will suffice, provided $a \geqq b$:

$$f_0(x) = \begin{cases} 0, & x < a \\ \dfrac{\sum_{k=1}^{n} \exp\left\{ -\frac{a}{2} k(k-1) \right\}}{\sum_{k=1}^{\infty} \exp\left\{ -\frac{a}{2} k(k-1) \right\}}, & na \leqq x < (n+1)a, \quad n \geqq 1; \end{cases} \tag{7a}$$

$$g_0(x) = \begin{cases} \exp\left\{ -\frac{1}{4a}(x^2 - \gamma^2), \right. & x < -\gamma, \quad \gamma = \dfrac{2a \log 2}{b} \\ 1, & x \geqq -\gamma. \end{cases} \tag{7b}$$

(The aforementioned restriction, $a \geqq b$, is immediately set aside by the following observation: If the unique solution which has been shown to exist for $a \geqq b$ is denoted as $f(x; a, b)$, then $1 - f(-x; b, a)$ satisfies (1) and the boundary conditions when $b > a$. Therefore, we may confine attention to the case $a \geqq b$.)

Thus, we have demonstrated the existence of a solution to the given problem, and it follows from (7a) and (7b) that the boundary values are approached with Gaussian rapidity:

$$f(x) = O(\exp\{-x^2/4a\}), \qquad x \to -\infty, \tag{8a}$$

$$f(x) = 1 - O(\exp\{-1 - \varepsilon)x^2/2a\}), \quad x \to +\infty. \tag{8b}$$

3. We now proceed to obtain an analytical expression for the solution whose existence has been demonstrated in the preceding section. The rapid approach of $f(x)$ to its limiting values at $\pm\infty$, as indicated by (8a) and (8b), guarantees the existence, for all values of the complex parameter s $(= \sigma + it)$, of the bilateral Laplace transform

$$\int_{-\infty}^{\infty} e^{-sx}\{f(x) - u(x)\}\, dx. \tag{9}$$

[Here $u(x)$ denotes the "unit step-function" $1/2(1+\operatorname{sgn} x)$.] Therefore, the integral

$$\tilde{f}(s) = \int_{-\infty}^{\infty} f(x)e^{-sx}\,dx \tag{10}$$

exists for $\sigma > 0$ and may be analytically continued to the whole (finite) s-plane except for a simple pole of residue $+1$ at the origin. From (1) one readily obtains, for $\sigma > 0$,

$$\tilde{f}(s)\{e^{as}-1\} = \tilde{f}(s+1)\{1-e^{-b(s+1)}\},$$

or

$$\tilde{f}(s) = \tilde{f}(s+1)e^{-as}\,\frac{1-e^{-b(s+1)}}{1-e^{-as}}. \tag{11}$$

To eliminate the factor e^{-as}, let

$$\tilde{f}(s) = \exp\left\{\frac{a}{2}(s^2-s)\}\right\} g(x). \tag{12}$$

Then from (11) and (12) one obtains

$$g(s) = g(s+1)\frac{1-e^{-b(s+1)}}{1-e^{-as}},$$

and more generally, for any positive integer k,

$$g(s+k-1) = g(s+k)\,\frac{1-e^{-b(s+k)}}{1-e^{-a(s+k-1)}}. \tag{13}$$

Writing out (13) for $k = 1,2,\cdots,n$ and then multiplying and simplifying, one obtains

$$g(s) = g(s+n)\,\frac{1-e^{-a(s+n)}}{1-e^{-as}}\,\frac{\prod_{k=1}^{n}\{1-e^{-b(s+k)}\}}{\prod_{k=1}^{n}\{1-e^{-a(s+k)}\}}. \tag{14}$$

Letting $n \to \infty$, one easily sees that the two products appearing in (14) converge for *all* s, not only for $\sigma > 0$; it follows that $g(s+n)$ converges to an entire function, which will be denoted as $h(s)$. Since $h(s) = \lim_{n\to\infty} g(s+n) = \lim_{n\to\infty} g(s+(n+1)) = \lim_{n\to\infty} g((s+1)+n)$, it follows that

(15) $$h(s+1) = h(s).$$

Thus, from (14) one obtains

(16) $$g(s) = \frac{h(s)}{1-e^{-s}} \frac{\prod_{k=1}^{\infty}\{1-e^{-b(s+k)}\}}{\prod_{k=1}^{\infty}\{1-e^{-a(s+k)}\}}$$

and hence

(17) $$f(s) = \frac{e^{(a/2)s^2}h(s)}{2\sinh as/2} \frac{\prod_{k=1}^{\infty}\{1-e^{-b(s+k)}\}}{\prod_{k=1}^{\infty}\{1-e^{-a(s+k)}\}};$$

Now, the infinite product appearing in the numerator in (17), which we shall henceforth denote by $b(s)$, possesses simple zeroes at the points $s = -k + (2m\pi i/b)$ $(k > 0, m \gtreqless 0)$, while the other product, which will be denoted $a(s)$, has simple zeros at $s = -k + (2m\pi i/a)$ $(k > 0, m \gtreqless 0)$. Thus, the quotient of these products is analytic and different from zero at the negative integers, but possesses simple poles at the points $s = -k + (2m\pi i/a)$ $(k > 0, m \neq 0)$. (Actually, this assertion is justified only if b/a is irrational, for otherwise some of the non-real zeroes of the numerator and denominator cancel each other out; nevertheless, the final result is easily seen to be valid for rational as well as irrational values of b/a.) Since $\tilde{f}(s)$ must be analytic everywhere except at the origin, the factor $h(s)$ must possess zeroes at the non-real zeroes of $a(s)$ and of $\sinh a_2/2$; i.e., at the points $-k + 2m\pi i/a$ $(k \geqq 0, m \neq 0)$. From the periodicity property (15) it then follows that $h(s)$ must vanish at the points $-k + (2m\pi i/a)$ $(k \gtreqless 0, m \neq 0)$. Now, the theta-function $\theta_1(\pi s, \exp(-2\pi^2/a))$, which for brevity will henceforth be denoted $\theta(s)$, is entire and possesses simple zeroes at the points $-k + 2m\pi i/a$ $(k \gtreqless 0, m \gtreqless 0)$; furthermore,

(18) $$\theta(s+1) = -\theta(s).$$

Hence, the entire function $\theta(s)/\sin \pi s$ has simple zeroes at precisely the points which have been shown to be zeros of $h(s)$, and it is periodic with period one. It follows that

(19) $$h(s) = \frac{\theta(s)}{\sin \pi s} H(s),$$

where $H(s)$ is also an entire function with period one. From (17) we now obtain

(20) $$H(s) = \frac{a(s)}{b(s)} \frac{2\sinh as/2}{\exp(as^2/2)} \frac{\sin \pi s}{\theta(s)} \tilde{f}(s).$$

It will be shown next that $H(s)$ must reduce to a constant. For all values of s, the inequality

(21a) $$\left| \frac{2\sinh as/2 \sin \pi s}{\exp(as^2/2)} \right| \leqq 2\exp\left\{ \frac{a|\sigma|}{2} - a\sigma^2 + \frac{at^2}{2} + \pi|t| \right\}$$

is easily verified; furthermore, for some positive integer N, the inequality

(21b) $$\left| \frac{a(s)}{b(s)} \right| < 2$$

holds whenever $\sigma \geqq N$, since each of the functions $a(s)$, $b(s)$ is readily seen to approach unity as $\sigma \to +\infty$. Now, to obtain a suitable estimate on $|\theta(s)|$, one employs the identity

(22) $$\theta\left(s + \frac{2\pi i}{a}\right) = -\theta(s)\exp\left(-2\pi i s + \frac{2\pi^2}{a}\right).$$

Using (22) repeatedly, one obtains for any positive integer n the equality

(23) $$\theta\left(s + \frac{2\pi n i}{a}\right) = (-1)^n \theta(s)\exp\left(-2\pi i n s + \frac{2\pi^2 n^2}{a}\right).$$

Thus, by setting $s = \sigma + (\pi i/a)$, one obtains the following inequality, which is valid everywhere on the line $t = (\pi(2n+1)/a$:

(24) $$|\theta(s)| \geqq C\exp\left\{\frac{2\pi^2 n + 2\pi^2 n^2}{a}\right\},$$

where C denotes the *positive* minimum of $|\theta(s)|$ on the line $t = \pi/a$. Finally, since $f(x)$ is everywhere positive, it follows from (10) that the inequality

(25) $$|\tilde{f}(s)| \leqq \tilde{f}(\sigma)$$

holds everywhere in the half-plane $\sigma > 0$. Taking account of these several inequalities, one finds that, on the line segment $N \leqq \sigma \leqq N+1$, $t = (\pi(2n+1)/a)$, the inequality

(26) $$|H(s)| \leqq C' \exp\frac{2\pi^2 n}{a}$$

holds, where $C' = (4/C)\exp(3\pi^2/2a)$. By periodicity, however, the above restriction $N \leqq \sigma \leqq N+1$ may be dropped, so that (26) holds everywhere on the line $t = \pi(2n+1)/a$. Now let s be confined momentarily to real values. As a smooth periodic function, $H(s)$ can be expanded in a convergent Fourier series:

$$H(s) = \sum_{k=-\infty}^{\infty} c_k e^{2\pi iks}, \quad c_k = \int_0^1 H(s)e^{-2\pi iks}\,ds.$$

Employing the Cauchy integral theorem, one obtains

$$c_k = \int_{\Gamma} H(s)e^{-2\pi iks}\,ds, \tag{28}$$

where Γ_n denotes the path consisting of the vertical and upper sides of the rectangle with vertices at 0, 1, $\pi(2n+1)i/a$, and $1 + (\pi(2n+1)i/a)$. By periodicity, the integrals along the vertical sides cancel, and one is left with the integral along the upper side of the rectangle. Taking account of (26), one obtains

$$|c_k| \leqq C' \exp\left\{\frac{2\pi^2}{a}[k + n(2k+1)]\right\}. \tag{29}$$

Since n may be chosen arbitrarily large, it follows from (2) that c_k must vanish for negative values of k. For positive values of k the same result is obtained by integrating in the lower half-plane. Therefore, $H(s)$ must reduce to a constant on the real axis, and hence, as asserted above, everywhere. The value of this constant is determined by the condition, stated after (10), that $\tilde{f}(s)$ must have a residue of $+1$ at the origin. In this way we are led to the formula

$$\tilde{f}(s) = \frac{C'' \exp as^2/2}{2\sinh as/2} \frac{\theta(s)}{\sin \pi s} \frac{b(s)}{a(s)}, \tag{30}$$

where

$$C'' = \frac{\pi a}{\theta'(0)} \frac{a(0)}{b(0)}. \tag{31}$$

Thus, the function $f(x)$ whose existence was demonstrated in the previous section must admit the integral representation

$$f(x) = \frac{1}{2\pi i}\int_{\rho_0 - i\infty}^{\rho_0 + i\infty} e^{sx}\tilde{f}(s)ds, \qquad \sigma_0 > 0. \tag{32}$$

Alternatively, one could have obtained (32), independently of the considerations of the preceding section, simply by assuming the existence of a solution to the given problem whose behavior near $\pm\infty$ permits the use of the analytic device which we have employed. It is then quite simple to justify this procedure *a posteriori* by showing directly that (32) defines a continuous function $f(x)$ satisfying (1) and the prescribed boundary conditions; as might be expected, the proof that the boundary conditions are satisfied involves a suitable change in the path of integration and an application of the Riemann-Lebesgue lemma. (In fact, $f(x)$ is easily shown to be analytic in the *complex* variable x in a strip of width 2π about the real axis, and to depend continuously on the parameters a and b.) However, we have been unable to establish directly from (32) either the monotonicity of $f(x)$ (or even the inequalities $0 \leqq f(x) \leqq 1$) or the estimates (8a), (8b).

4. It is to be expected that the case $a = b$ should be particularly simple. By a quite elementary argument one obtains the solution [2, p. 106]

$$f(x) = \frac{\sum_{n=0}^{\infty} \exp\left\{-\frac{(x-na)^2}{2a} + \frac{x-na}{2}\right\}}{\sum_{n=-\infty}^{\infty} \exp\left\{-\frac{(x-na)^2}{2a} + \frac{x-na}{2}\right\}}. \tag{33}$$

It may be remarked that no assumption of any sort, even of boundedness or measurability, is needed. It might be of interest to investigate whether uniqueness can be demonstrated in the case $a \neq b$ under conditions appreciably weaker than boundedness.

Finally, we remark that by taking account in (30) and (32) of the expansion

$$\theta(s) = \text{constant} \cdot \sum_{k=0}^{\infty} (-1)^k \exp\{-2\pi^2 k(k+1)/a\} \sin(2k+1)\pi s \tag{34}$$

we may obtain, in the case $a \neq b$, a series expansion of $f(x)$ which constitutes a generalization of (33); the rapid convergence of (34) will presumably be reflected, for "reasonable" values of a and b, in the rapid convergence of the expansion of $f(x)$.

5. I wish to express appreciation to Professor N. J. Fine, with whom I have discussed this problem. His solution, worked out quite independently, has some overlap with the one presented here.

REFERENCES

1. L. Kanal, Psychometrika, **27**, (1) (1962), 89–104.
2. L. Kanal, ibid, 105–109.

UNIVERSITY OF NEW MEXICO
AND
OFFICE OF NAVAL RESEARCH