On one-sided smoothing of event sequences.

We consider a finite number of events E_i (i from 0 through N) occurring at moments t_i ($i>j \Rightarrow t_i > t_j$). The problem I found myself faced with was how to define a continuous function $f(t)$ that could be interpreted as "the event frequency at moment t." (I encountered this problem when trying to discover a strategy for deciding how much primary store should be allocated to independent programs in a multiprogrammed environment when for each program a "target page fault frequency" had been decided.)

A possible definition of the event frequency is in terms of the Dirac function:

$$f(t) = \sum_{i=0}^{N} \delta(t - t_i) \qquad (1)$$

which has the obvious advantage that it satisfies

$$\int_{-\infty}^{+\infty} f(t).\, dt = N+1 \qquad (2)$$

but the equally obvious disadvantage of severe discontinuities, which makes the comparison of $f_0(t)$ and $f_1(t)$ — two different frequencies corresponding to different values of the parameter I could control — rather difficult. Hence our desire to smooth the data. The smoothing, however, had to be one-sided in the sense that the smoothed value $f(t)$ may at any moment t only depend on the set of values t_i with $t_i \leq t$. In order to make a sensible choice among the wealth of possibilities for $f(t)$ we should impose upon $f(t)$ as many "sensible" constraints as we can think of. To start with, property (2) seems a good choice. The next one is that we should realize that frequencies — as usually understood — have a dimension "$time^{-1}$", i.e. if we consider a next sequence of events E_i', occurring at moments $t_i' = at_i$ and define the corresponding frequency $f'(t)$ for that second event sequence, then

$$f'(at) = \frac{1}{a} f(t) \qquad (3)$$

should hold. Property (3) is not in conflict with requirement (2), for

$$\int_{-\infty}^{+\infty} f'(t')\,dt' = \int_{-\infty}^{+\infty} f'(at)\,d(at) = \int_{-\infty}^{+\infty} a.f'(at).dt =$$

$$\int_{-\infty}^{+\infty} f(t).dt = N+1$$

Reasonable as requirement (3) may seem, it is (for finite sequences) too much to hope for: if $t_0 = 0$, the relation (3) will not hold for $t_0 \leq t \leq t_1$, for how are we going to guess "a"? Therefore, we must loosen it and can only require (3) to hold asymptotically for $t > t_i$ with sufficiently large i.

Relation (2), however, still stands rather firmly. Because in (2) we have only used

$$\int_{-\infty}^{+\infty} \delta(t)\, dt = 1$$

we could consider functions $w_i(t)$ such that

$$w(t) = 0 \quad \text{for} \quad t < 0 \tag{4}$$

and

$$\int_0^{\infty} w(t).\, dt = 1 \tag{5}$$

and replace (1) by

$$f(t) = \sum_{i=0}^{N} w_i(t - t_i) \tag{6}$$

where (4) does credit to the one-sidedness of the required smoothing, (5) guarantees (2) and (6) makes the whole definition invariant for shifting of the origin. The question is whether we can choose the w_i in such a way — satisfying (4) and (5) — as to approximate (3) at least asymptotically.

A reasonable first guess — corresponding to fading out of the past in an exponential way — for $w_i(t)$ is with $k_i > 0$:

$$\begin{aligned} w_i(t) &= 0 && \text{for} \quad t < 0 \\ &= k_i . e^{-k_i . t} && \text{for} \quad t \geq 0 \end{aligned} \tag{7}$$

where we can try to adapt k_i — on account of past history, if any — towards a better approximation of (3). I called (7) a first guess, because we also wanted a continuous function $f(t)$ and with (6) and (7), $f(t)$ is discontinuous for any $t = t_i$. A continuous f would follow from the proper linear composition — proper in view of (4), (5) and $w_i(t) = 0$ —

$$w_i(t) = 0 \qquad \text{for } t < 0$$
$$= \left(e^{-k_i \cdot t} - e^{-h_i \cdot t}\right) \cdot \frac{h_i \cdot k_i}{h_i - k_i} \qquad \text{for } t \geq 0 \qquad (8)$$

and now we are free in the choice of both h_i and k_i, provided that they are both positive and different from each other. I am severely tempted to restrict myself to

$$h_i = c * k_i \qquad \text{with} \quad c \neq 1 \text{ and } c > 0 \qquad (9)$$

the reason being that both h_i and k_i are of dimension "$time^{-1}$" and that I cannot forget my target (3).

*   *

*

In order to come to grips with the constant c I studied (omitting subscripts "i") according to (8)&(9)

$$w(t) = k \cdot \frac{c}{c-1} \cdot \left(e^{-kt} - e^{-ckt}\right)$$

and tried to determine c so as to minimize the maximum value of

$$\frac{c}{c-1}\left(e^{-kt} - e^{-ckt}\right)$$

(I wanted my ripples as smooth as possible.) As this led to the inacceptable value $c = 1$, I took the limit for $c \to 1$ and found — and this is my final suggestion —

$$\begin{aligned} w_i(t) &= 0 && \text{for } t < 0 \\ &= k_i^2 t \cdot e^{-k_i t} && \text{for } t \geq 0 \end{aligned} \qquad (10)$$

Also (10) has the property that $\int_0^\infty w_i(t)\,dt = 1$.

*   *
  *

The next thing to decide is, how to choose the values k_i. A constant value for k_i seems in view of (3) out of the question, because k_i has the dimension of a frequency. Apart from initial difficulties it seems reasonable to choose k_i proportional to $f(t_i)$ i.e.

$$k_i = d * f(t_i) \qquad (11)$$

To get some feeling for a reasonable value for d, we consider $f(t)$ for $t \geq 0$ as results from an infinite sequence of events that have occurred at $t = 0, -1, -2, -3, \ldots\ldots$ etc; for all passed events we may assume the same value $k_i = k$ and we find for $f(t)$, according to (6) and (10)

$$f(t) = k \cdot \sum_{i=0}^{\infty} k \cdot (t+i) \cdot e^{-k \cdot (t+i)}$$

which, indeed, satisfies $f(0) = f(1)$ as was to be expected.

Note. I summed the sequence and found

$$f(t) = \left(\frac{k^2 \cdot e^{-k \cdot t}}{1 - e^{-k}} \right) \cdot \left(t + \frac{e^{-k}}{1 - e^{-k}} \right)$$

(End of note.)

The functions (10) have all one maximum, viz. at $t = k_i^{-1}$. We would like to attribute the maximum of $f(t)$ for $0 \leq t \leq 1$ to the term with $i = 0$ (i.e. the last event). For $i > 0$ all the terms should be decreasing functions of t for $t > 0$. We can expect the smoothest maximum of $f(t)$ if it is in the neighbourhood of $t = \frac{1}{2}$ and this suggests k in the neighbourhood of 2 and

$$d = 2 \qquad (12)$$

For, because in the average $f(t) = 1$, we know that $f(0) = f(1) < 1$, therefore with $d = 2$ we shall find $k < 2$ and the term with $i = 0$ will reach its maximum at $t_0 > \frac{1}{2}$. Because the contribution of the remaining terms is a decreasing function of t, this will be compensated and $f(t)$ will find its maximum at $t_f < t_0$.

Note. Using I sliderule I have investigated the consequences of the choice $d=2$. Salvo errore et omissione I found for $k = 1.616$, $f(0) = f(1) = 0.808$ as minimum value for $f(t)$ and $f(0.41) = 1.105$ as maximum value. Anyone that would like to rely on these figures should reestablish them in a more reliable manner. I only mention these results because they represent the outcome of an hour's work and give some impression of how smooth $f(t)$ becomes in the case of equally spaced events. If it were of importance, I would try slightly smaller values of d as well. (End of note.)

19th February 1975
Plataanstraat 5
NUENEN - 4565
The Netherlands

prof. dr. Edsger W. Dijkstra
Burroughs Research Fellow