Temporal Geometry in Topotime

Karl Grossner and Elijah Meeks, Stanford University Libraries

It is commonplace to think about timespans in geometric terms and to draw timelines representing the timespans of events (Rosenberg & Grafton 2012). A timespan is often drawn as a straight line, possibly having dashed or otherwise differentiated segments at either end, indicating uncertainty about its extents (Fig 1). When depicting the sub-event parts of a larger event or period, we may widen the line to a rectangle (Fig 2). The *spatialization* of non-spatial things in this way and many others is a routine practice in human cognition, visualization, and computation (cf. Skupin & Fabrikant 2008; also Lakoff & Johnson 1980; Johnson 1987 on spatial metaphor).

Figure 1 – A timespan with uncertain start and end

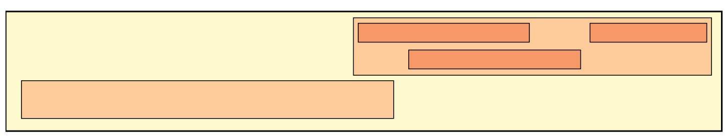

Figure 2 – A composite event rendered as its timespan with two levels of sub-event parts.

One way of representing the uncertain bounds of timespans computationally is with a *quad* of dates or date parts; examples include Simile Timeline's [*start*, *latestStart*, *earliestEnd*, *end*], and Topotime has adopted that terminology. These correspond to [*fuzzybegin*, *begin*, *end*, *fuzzyend*] used in (Kaupinnen, et al 2010). For Topotime the outer terms, *start*/*end* have the meaning "not before" and "not after." The terms *terminus post quem* and *terminus ante quem*, although in common usage, can imply "for which there is evidence," whereas in practice temporal assertions can be estimates, and material evidence either absent or uncited.

For Topotime we have followed Kaupinnen, et al (2010) in characterizing timespans as a probability functions having one to three principal segments, and reasoning about their curves and the areas under those curves (AUC). Y-axis values of between 0 and 1 can be considered as probabilities or alternatively, 'percent certainty.' Where timespan (tSpan) segments for *start*/*latestStart* or *earliestEnd*/*end* exist, they can be articulated by interim points with or without smoothing curves.

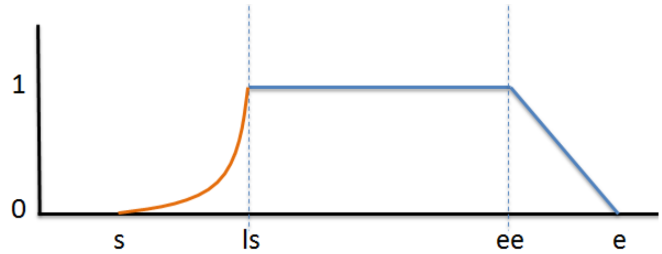

Figure 3 – A timespan (tSpan) as probability function, rendered as a curve.

Mathematical functions are not geometrical, but we routinely use lines, polylines, curves, and areas to reason about them visually and spatially. As Kaupinnen has shown, we can close the AUC for a timespan function along the x-axis, and calculate its area geometrically or as an integral. This prompts us to ask what the meaning of that AUC is; i.e. what is the calculated area value a sum of? For Topotime, it is something like an "asserted certainty," explained as follows. The X-axis represents Julian days and the Y-axis, either probability, certainty, confidence, or likelihood, however conceived by the investigator.

Consider a Period with the following tSpan:

```
[{s: 1744-09-30, ls: 1745-01-01, ee: 1745-12-31, e: 1746-06-30,
      eee: [[0.2,0.35],[0.7,0.20],[0.9,0.08]]}]
```

This event certainly occurred throughout 1745, may have begun three months prior, and may have ended six months afterward, with certainty diminishing as described by the "eee" term. In "eee" coordinate pairs, the first term is a percentage of the distance between "ee" and "e" and the second, *y* for that *x*.

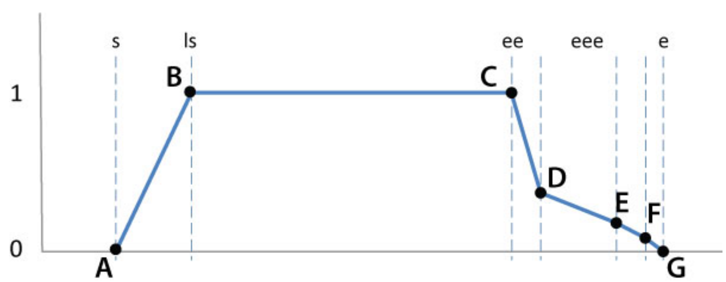

Figure 4 – curve of above tSpan

Among the data objects generated for this tSpan by one of the Topotime parsers, *periodic.py*, is the polygon ABCDEFG, as shown in Listing 1 and Figure 4. Gregorian dates are converted to Julian dates (÷ 10^6 here) making for these coordinates:

```
{'x': 2.358315, 'y': 0}, {'x': 2.358408, 'y': 1},
{'x': 2.358772, 'y': 1}, {'x': 2.3588082, 'y': 0.35},
{'x': 2.3588987, 'y': 0.2}, {'x': 2.3589349, 'y': 0.08},
{'x': 2.358953, 'y': 0}, {'x': 2.358315, 'y': 0}]
```

Listing 1 – temporal geometry of above tSpan

Given such geometries, and collections of them, we can ask and answer several questions.

- For any value or range of values on the X-axis (i.e. dates in this example), we can return a likelihood *L* that period *P* was true. For a single value, *L* will equal *y(x)*. For a set (x_i, ..., x_n) a mean of (y_i, ..., y_n) can be useful.
- For any two tSpans in the Collection we can calculate an area of intersection (overlap); this value is meaningful as a relative level of temporal coincidence, as demonstrated by Kaupinnen (2010). (See Figure 6).
- We can provide a query timespan and return a list of the periods in the collection that overlap with it, ordered by the area of intersection. This is demonstrated in the exemplar interface at http://dh.stanford.edu/topotime/demo_py.html
- Potentially, we can calculate the probability of two events overlapping. For example, of two individuals having met as adults if we model their adulthoods as sub-periods in their lifespans.

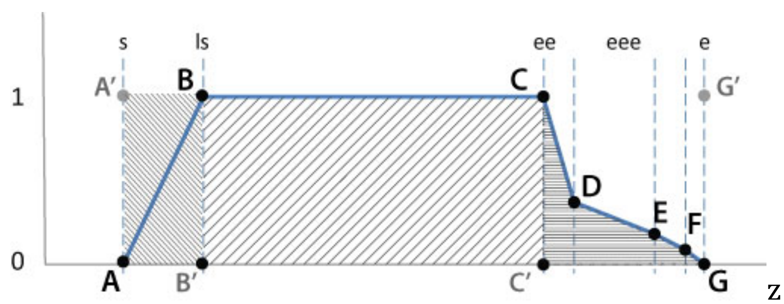

Figure 5 – Percentages of "potential certainty"

In Figure 5, the area of region B'BCC' is 365.25 units (days in a year); we could say the "potential certainty" for that interval is 365.25, and that the certainty of a day-long event between B' and C' being true is 1.

The area of region ABB' is 45, one-half of the "potential certainty" for the 90-day span represented by AA'BB'. Therefore the certainty for that period during that interval is on average 0.5; for a given *x*, it is *y*. The potential certainty area of region C'CG'G is 180 units (days here), and the area under curve (the polygon C'CDEFG) can be computed as trapezoids. Topotime development plans include generating smooth curves from points, the specification of functions in tSpan "sls" and "eee" terms, and computing AUCs as integrals in those cases.

Figure 6 below shows three overlapping tSpans and a query indicated by a rectangle. Period C coincides more with Period B than with Period A, according to the estimations of their author. The query, "What was true during timespan Q?" returns the ordered list, [B, A, C], based on calculations of percent overlap. Note that this does not (yet) tell us what the probability of any union of these was. The area within intersections has meaning only relative to other intersections. We plan to develop more ways of reasoning about these temporal geometries in the near future.

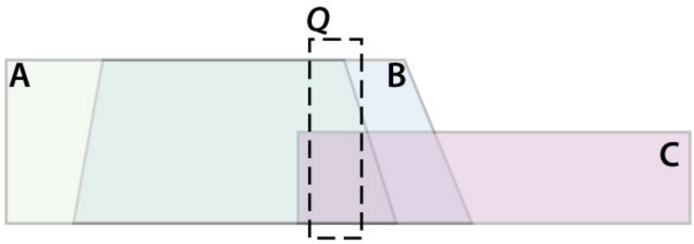

Figure 6 – Three overlapping periods in a collection and a query box; C coincides with B more than with A.

Works cited

Johnson, M. (1987). *The body in the mind: The bodily basis of meaning, imagination and reason*. Chicago: University of Chicago Press.

Kauppinen, T. Mantegari, G., Paakkarinen, P., Kuittinen, H., Hyvonen, E., Bandini, S. (2010). Determining relevance of imprecise temporal intervals for cultural heritage information retrieval. *International Journal of Human-Computer Studies* 68 (2010) 549–560

Lakoff, G. and M. Johnson (1980). *Metaphors we live by*. Chicago: University of Chicago Press.

Rosenberg, D. and Grafton, A. (2012). *Cartographies of Time*. Princeton, NJ: Princeton Architectural Press

Skupin, A. and Fabrikant, S. (2008) Spatialization. in: Wilson, J., and Fotheringham, S. (Eds.) ***The Handbook of Geographical Information Science***. Blackwell Publishing. 61-79.