A generalization of the Lexis surface

Jonas Schöley

March 31, 2017

The Lexis surface[1] is a tool for working with time scales which are linearly dependent on each other, e.g. the two points in time "current period" and "birth cohort" are related by their difference – the "current age". This relationship can be generalized to n points in time and the m durations between any pair of points. We show that the Lexis diagram, the Lexis' marriage diagram and the APCTDL space are instances of such a generalized Lexis diagram. We start by formalizing the notions of points in time $\mathbf{p}$ and durations $\mathbf{d}$ between these points.

Definition 1: Let $\mathbf{p}$ be a vector with n elements where $n \geq 2$ and elements $p_{i=1} \dots p_n \in \mathbb{R}$. Let $\mathbf{d}$ be a vector with m elements such that $\mathbf{d}_{ij} = p_i - p_j$ for all i and $j = 1, \dots, n$ where $i > j$.[2]

Theorem 1: The number of elements in $\mathbf{d}$ is $m = n(n-1)/2$.

Proof: Because d_{ij} must satisfy $i > j$ it can be thought of as all the elements below the subdiagonal of a $n \times n$ matrix of which there are $m = n(n-1)/2$.

Corollary 1.1: The sum of the number of elements in $\mathbf{p}$ and $\mathbf{d}$ is the number of edges on a $n+1$-polytype.

Proof: The sum of the number of elements in p and d is $n + m = n + n(n-1)/2$ which is the number of edges on an $n+1$-polytype.

Instead of a particular choice of n-points in time we now consider the set P of all possible choices of n points in time.

Definition 2: Let $P = \mathbb{R}^n$ be the standard basis vector-space spanning all $\mathbf{p}$.

Theorem 2: The linear transformation $T_{\mathbf{A}} : P \to D$ where $D = \mathbb{R}^{n(n-1)/2}$ and $\mathbf{A}_T\mathbf{p} = \mathbf{d}$ is given by[3]

$$\underbrace{\left[\begin{array}{c|c|c} -\mathbf{I}_1 & \mathbf{1}_1 & \mathbf{0}_{1\times n-2} \\ -\mathbf{I}_2 & \mathbf{1}_2 & \mathbf{0}_{2\times n-3} \\ \vdots & \vdots & \vdots \\ -\mathbf{I}_{n-1} & \mathbf{1}_{n-1} & \left(\mathbf{0}_{n-1\times 0}\right) \end{array}\right]}_{\mathbf{A}_T} \times \underbrace{\begin{bmatrix} p_1 \\ \vdots \\ p_n \end{bmatrix}}_{\mathbf{p}} = \underbrace{\begin{bmatrix} d_1 \\ \vdots \\ d_{n(n-1)/2} \end{bmatrix}}_{\mathbf{d}}.$$

Given *any* n points in time it is always possible to calculate the durations between any pair of these points. This calculation can be understood as a linear transformation of the *point-space* P to the *duration-space* D.

Proof: A_T is a simple consequence of the definition of $\mathbf{d}$.

Corollary 2.1: There exists no linear map $T_{\mathbf{A}}^{-1} : D \to P$.

[1] We understand the "Lexis-surface" as the Lexis diagram without the notion of life-lines. While life-lines may be added to the generalized Lexis surface they are not part of our definition – e.g. a period-age Lexis surface featuring a life-line with clearly marked "death cohort" does not constitute an extension of the Lexis surface by death-cohort. Such an extension is only achieved by adding an orthogonal axis to the period-age plane featuring all possible death-cohorts.

[2] Equivalently $\mathbf{d}$ can be expressed as $\text{vech}(\mathbf{p} \times \mathbf{1}_n^{\mathsf{T}} - \mathbf{1}_n \times \mathbf{p}^{\mathsf{T}})$. For ease of notation we use double indices to index each element of the vector $\mathbf{d}$. The indices ij of the k^{th} element of $\mathbf{d}$ are given by $d_{ij} = d_k = p_{i=\lfloor \frac{1}{2}+\sqrt{2k} \rfloor + 1} - p_{j=k-C\left(\left\lfloor \frac{1+\sqrt{8k}}{2} \right\rfloor\right)}$ with $C[f(k)]$ being the binomial coefficient $\binom{f(k)}{2}$.

[3] $\mathbf{I}_n$ is the $n \times n$ identity matrix; $\mathbf{1}_n$ is a vector with n elements where each element is 1; $\mathbf{0}_{m\times n}$ is a $m \times n$ matrix where each element is 0. For notational convenience we allow for a matrix with 0 columns, written $\left(\mathbf{0}_{n\times 0}\right)$.

Given only a set of m durations it is impossible to identify a single set of n points in time marking the endpoints of the durations, e.g. given only age one can not derive birth cohort and period.

Proof: $T_{\mathbf{A}}^{-1}$ exists if and only if $\mathbf{A}_T^{-1}$ exists. $\mathbf{A}_T$ has no inverse since the columns of $\mathbf{A}_T$ are always linearly dependent on each other, i.e. the last element in each row of $\mathbf{A}_T$ is always the negation of the sum of all the other row elements. Therefore $T_{\mathbf{A}}^{-1}$ does not exist.

Theorem 3: *(Work in progess)* Something that states that you can transform $P^n \leftrightarrow M^n$ where the basis vectors of M are a mixture of point dimensions and duration dimensions.

Theorem 4: *(Work in progess)* There are $b =?$ many ways to choose n elements out of $\{p_{i=1,\dots,n}, d_{k=1,\dots,m}\}$ whoose linear combination yields one or more remaining elements of $\{p_{i=1,\dots,n}, d_{k=1,\dots,m}\}$.

Example 1: The Lexis surface

Let $\mathbf{p}$ have two elements. Then, following definition 1, $\mathbf{d}$ is defined as

$$d_1 = p_2 - p_1.$$

Interpreting d_1 as *age*, p_2 as *period* and p_1 as *birth cohort* yields the classic APC identities. The Lexis surface as the plane of all possible combinations between *age* and *period* with cohorts as diagonals results from the transformation $P^2 \to M^2$ as shown to be possible in theorem 3.

Example 2: Lexis' marriage identity

Along with his well known 2-dimensional diagram Lexis (1875) also published a 3-dimensional extension which he applies to marriage and separation processes.

Let $\mathbf{p}$ have three elements. Then $\mathbf{d}$ is defined as

$$\begin{aligned} d_1 &= p_2 - p_1 \\ d_2 &= p_3 - p_1. \\ d_3 &= p_3 - p_2 \end{aligned}$$

Interpreting p_1 as *birth cohort*, p_2 as *marriage cohort* and p_3 as *separation cohort* yields the durations *age at marriage* (d_1), *age at separation* (d_2), and *duration of marriage* (d_3). Lexis' "marriage space" is reconstructed by transforming $P^3 \to M^3$ where M has three orthogonal basis vectors corresponding to (p_1, d_1, d_3).

Example 3: Adding death cohort to the Lexis surface

Equivalent to example 2 we start with a three element vector $\mathbf{p}$ yielding the very same identities as shown above.

Interpreting p_1 as *birth cohort*, p_2 as *period* and p_3 as *death cohort* yields the durations *age* (d_1), *lifespan* (d_2), and *time until death* (d_3). This vector space P^3 contains the Lexis surface as a sub-space.