Counting kin in an age×stage structured network

Joe William Barrass Butterick[1], Joanne Ellison[1], Peter. W. Smith[1], Erengeul Dodd[1], Jakub Bijak[1], Jonathan J. Forster[2], and Jason Hilton[1]

[1]University of Southampton
[2]University of Warwick

1   Introduction

Mathematical models of kinship stem from the influential papers of Goodman et al. (1974) and Pullum (1982). More recently, Caswell (2019) applied a system of matrix projections, one for each kin of some random "Focal" population member. Subsequent to its inception, the matrix based framework has been extended to age×stage (multi-state) structured populations (Caswell, 2020). Developing the theoretical synthesis of Caswell (2020) to cover arbitrary stage, however, presents challenges. A paper currently in development (Roper & Caswell, 2023) illustrates the complexities of considering spatial stage. In this research we extend age×stage kinship while drawing from a different mathematical perspective. Our approach draws from Markov theory to project kin backwards and forwards in time; move through family lineages; and calculate cross-generational expected age×stage distributions of kin. Similar methods have been applied to model stage based demography (Coste et al., 2021). We derive three concise formulae to describe individuals' age×stage kinship networks. We illustrate our approach by defining stages as spatial locations, exemplified by clusters of local authority districts (LADs) taken from the National Records of Scotland (NRS) data. Future work will investigate an extension of the presently developed framework to account for time-varying demographic rates, and fit them using new Office for National Statistics (ONS) England and Wales data.

2   Methods

2.1   Notation

Let $M_{n,m}(\mathbb{R})$ denote the set of $n \times m$ matrices over the real numbers. Vectors are written boldface, lowercase: $\mathbf{x} = (x_1, x_2, ..., x_n)^\dagger$ where $\dagger$ represents the adjoint (conjugate transpose); matrices boldface, uppercase. Block diagonal matrices are denoted $\mathbb{A}$. The Schur product, written $\circ$, takes element-wise products of vectors. The vec-operator, which stacks columns of a matrix $\mathbf{A} \in M_{n,m}$ into a vector $\mathbf{a} \in M_{nm,1}$, is written vec(.). The vec-commutation matrix $\mathbf{\Phi}_{q,r}$ is defined through $\mathbf{\Phi}_{q,r}\text{vec}\mathbf{A} = \text{vec}\mathbf{A}^\dagger$ for $\mathbf{A} \in M_{q,r}$. We write $||\mathbf{x}||_1 = \sum_i^n |x_i|$. The spectral radius $\rho(\mathbf{A})$ of a matrix $\mathbf{A} \in M_{n,n}$ is the maximum modulus of its eigenvalues.

2.2 The established age×stage kin method

The model of Caswell (2020) considers a population structure defined through stages $i = 1, .., s$ and ages $j = 1, ..., n$. As such, the matrix

$$\mathbf{K} = \begin{pmatrix} k_{11} & k_{12} & \dots & k_{1n} \\ k_{21} & k_{22} & \dots & k_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ k_{s1} & k_{s2} & \dots & k_{sn} \end{pmatrix} \in M_{s,n} \tag{1}$$

has scalar entries k_{ij} which give the number of kin in stage i, of age j. One proceeds to project a block structured kinship vector $\tilde{\mathbf{k}} = \mathrm{vec}(\mathbf{K})$, using the following vec-permutation approach (Magnus & Neudecker, 2019). For each age class $j \in \{1, 2, .., n\}$, one defines stage transfer matrices $\mathbf{T}_j = \left(t_{kl}^{(j)} \right)$ with entries capturing transitions of individuals from stage l to k; and fertility matrices $\mathbf{F}_j = \left(f_{kl}^{(j)} \right)$ with entries specifying the production of stage k individuals by stage l individuals. For each stage $i \in \{1, 2, ..., s\}$, one defines age-specific mortality matrices $\mathbf{U}_i = \left(u_{kl}^{(i)} \right)$ specifying survival probabilities upwards in age classes (i.e., from age l to k); and recombination matrices $\mathbf{H}_i = \left(h_{kl}^{(i)} \right)$ which, independently for each stage, characterise the age-classes k of newborns from mothers of age l.

Direct sums of the above matrices (e.g., $\mathbb{U} = \mathbf{U}_i \oplus \dots \oplus \mathbf{U}_s$) determine the projection matrix $\tilde{\mathbf{A}} = \tilde{\mathbf{U}} + \tilde{\mathbf{F}}$, where $\tilde{\mathbf{U}} = \boldsymbol{\Phi}_{s,n}^\dagger \mathbb{U} \boldsymbol{\Phi}_{s,n} \mathbb{T}$ and $\tilde{\mathbf{F}} = \boldsymbol{\Phi}_{s,n}^\dagger \mathbb{F} \boldsymbol{\Phi}_{s,n} \mathbb{H}$. The actions of $\tilde{\mathbf{U}}$ and $\tilde{\mathbf{F}}$ are well explained in Caswell et al. (2018) where we refer the reader for further details. The age× stage of a particular type of Focal's kin, $\tilde{\mathbf{k}}$, when Focal is $x + 1$, is given by $\tilde{\mathbf{k}}(x+1) = \tilde{\mathbf{U}}\tilde{\mathbf{k}}(x) + \tilde{\boldsymbol{\beta}}(x)$. The range of $\tilde{\mathbf{k}}$ extends the family tree illustrated in Fig 1. Caswell (2019).

2.3 An alternative Markov model of age×stage kinship

Following the approach of Pullum (1982), we let B be the minimum genealogical distance between Focal and a common ancestor relating it to its kin. Let A be the distance from the kin and the common ancestor (e.g., Focal's sisters are defined $A = B = 1$; and daughters $A = 1, B = 0$). Defining kin through A, B naturally appeals one to travel backwards in time and locate Focal's B^{th} ancestor; and then travel forwards in time, probabilistically counting production and survival Focal's B^{th} ancestor's A^{th} descendant. To do the above, we apply to the genealogical Markov chains (Tuljapurkar, 1982; Bienvenu et al., 2017) associated with the age×stage block structured projection matrices.

We assume a (time-invariant) irreducible age×stage projection matrix $\tilde{\mathbf{A}} = \left(\tilde{a}_{ij} \right)$ so that Perron-Frobenius theory (e.g., Th. 8.4.4, Horn & Johnson (2010)) guarantees a stable population structure $\tilde{\mathbf{v}}$ and growth rate $\lambda = \rho(\tilde{\mathbf{A}})$. As such, the probability that an individual transitions from age×stage l at time $t-1$ to age×stage k at t, is given by

$$\mathrm{Prob}(l \to k) = \frac{\tilde{a}_{kl}\tilde{v}_l}{\lambda \tilde{v}_k} =: \tilde{p}_{kl}. \tag{2}$$

The genealogical Markov chain $\tilde{\mathbf{P}} = \left(\tilde{p}_{kl} \right)$ has entries defined through Eq (2), i.e., the probabilities that individuals currently in age×stage k (at time t) directly come from age×stage l (at time $t-1$).

One can decompose $\tilde{\mathbf{P}} = \tilde{\mathbf{P}}_U + \tilde{\mathbf{P}}_F$. Here, $\tilde{\mathbf{P}}_U = \left(\tilde{p}_{kl}^{(U)} \right)$, $\tilde{p}_{kl}^{(U)} = \tilde{u}_{kl}\tilde{v}_l/(\lambda \tilde{v}_k)$ provides probabilities that individuals currently in age×stage k were extant one time step previous and in stage l. Whereas,

$\tilde{\mathbf{P}}_F = \left(\tilde{p}_{kl}^{(F)}\right)$, $\tilde{p}_{kl}^{(F)} = \tilde{f}_{kl}\tilde{v}_l/(\lambda\tilde{v}_k)$ yields probabilities that individuals currently in age×stage k were produced by an ancestor in age×stage l one time step prior.

Using the above decomposition of the genealogical Markov chain, we can calculate for any of Focal's kin, the probabilistic trajectory which led to its current age×stage starting at some prior time. All we require to project kin backwards and forwards, is some well defined initial condition for Focal: $\tilde{\boldsymbol{\phi}}_0^{\dagger}$. Focal's mother's age×stage distribution at birth of Focal, $\tilde{\boldsymbol{\phi}}_0^{\dagger}\tilde{\mathbf{P}}_F$, is used to project backwards and forwards over timescale τ, counting survival and reproductive events.

2.3.1   Example of our method: kin descending from younger sisters of Focal's $(B-1)^{th}$ ancestor

As example, consider some kin of Focal which are produced via a lineage descending from younger sisters of Focal's $(B-1)^{th}$ ancestor. For a graphical illustration, see the right branches in Fig 1., Caswell (2019). Our method to predict these types of A, B kin is to concomitantly go backwards in time and locate Focal's B^{th} ancestor at time of producing Focal's $(B-1)^{th}$ ancestor; and come forwards in time accounting for all potential production of Focal's B^{th} ancestor's A^{th} descendant. By locating Focal's B^{th} ancestor at birth of Focal's $(B-1)^{th}$ ancestor, our method guarantees that Focal's A, B kin descend from younger sisters of Focal's $(B-1)^{th}$ ancestor. For example, we may wish to consider Focal's aunts which are younger than Focal's mother. The probability distribution of Focal's grandmother when producing Focal's mother, conditional on Focal's mother living for T years prior to producing Focal, is given by $\left(\tilde{\boldsymbol{\phi}}_0^{\dagger}\tilde{\mathbf{P}}_F\tilde{\mathbf{P}}_U^T\tilde{\mathbf{P}}_F\right)^{\dagger}$. As grandmother produces mother $1+T$ years prior to Focal's birth, to calculate Focal's expected number of younger aunts when born, we simply count over all possible combinations of one reproduction $\tilde{\mathbf{F}}$ and T survivals $\tilde{\mathbf{U}}$ which span $1+T$ years forwards (up to Focal's birth time). In the full manuscript we present concise formulae for such kin, as well as kin of Focal which descend through an older sister of Focal's $(B-1)^{th}$ ancestor, and direct descendants of Focal.

3   Application

Geography is an interesting descriptor of stage. Focal's kin can be born into any spatial location and migrate between regions before or during Focal's life. As such, transitions which define the resultant stochastic stage transfer matrices prove more complex than the example of parity considered in Caswell (2020). A model under concurrent development (Roper & Caswell, 2023), and based on that framework applies additional assumptions to deal with this complexity. Here, we show how our model provides an alternative solution.

In order to inform the matrices defined in section 2.3, we use NRS data spanning 1991-2021, and consisting of female only age-specific population size at the mid-year. Fertility data consisting of 5-year age classes from $15-19$ to $40-44$, is combined with migration and mortality data, consisting of 5-year age classes $0-4$ to $90-94$ to provide a model input of 19 age-classes $j = 0-4, 5-9, 10-14, ..., 85-89, 90-94$, over 32 LADs in Scotland. We cluster age-specific fertility rates across the period to demacerate LADs into one of 4 geographic clusters (via a $K-$means algorithm) $i = K_1, K_2, K_3, K_4$.

Our framework predicts an expected age×stage distribution of Focal's younger sisters, as illustrated in Fig 1. We see that Focal has no younger sisters at birth, but has gained some when leaving age class 0-4. The largest increase in the expected number of younger sisters occurs during years 0-4 of Focal. Younger sisters most likely stay in the same geographic region as the one Focal was born into, but do display monotonic

increasing (albeit small) propensities to migrate up to Focal's age class of around 80-84. Around the age class 20-24, Focal has accumulated their maximum number of younger sisters.
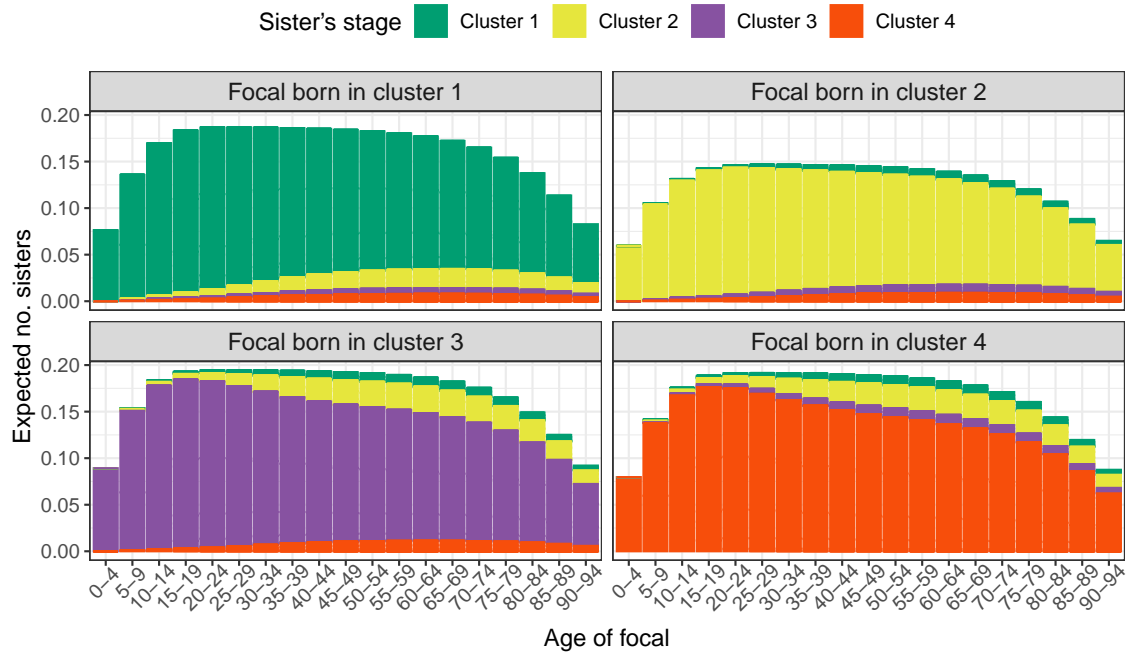

Figure 1: $A = 1, B = 1$: Spatial distribution of Focal's younger sisters as a function of Focal's age, x.

4   Discussion & outlook

This research presents a novel theoretical framework for age×stage structured kinship. Our method circumvents unwieldy theoretical calculations required when applying to the existing dominant framework (e.g., large systems of matrix projections). All we need to predict kin are well-defined initial age×stage distributions of Focal. In the sense that we avoid implementation of a large system of equations, our method provides computational benefits. Future research will extend the above model to incorporate time-dependent demographic rates. Applying to recently accessed England and Wales ONS data, the updated model will investigate temporal changes in age×geography of kin. We are presently engaged in developing these methods.

References

Bienvenu, F., Akay, E., Legendre, S., & McCandlish, D. M. (2017). The genealogical decomposition of a matrix population model with applications to the aggregation of stages. *Theor. Pop. Biol.*, *115*.

Caswell, H. (2019). The formal demography of kinship: A matrix formulation. *Demographic Res.*, *41*, 679–712.

Caswell, H. (2020). The formal demography of kinship II: Multistate models, parity, and sibship. *Demographic Res.*, *42*(38).

Caswell, H., de Vries, C., Hartemink, N., Roth, G., & van Daalen, S. F. (2018). Age stage-classified demographic analysis: a comprehensive approach. *Ecol. Monogr.*, *88*(4).

Coste, C. F. D., Bienvenu, F., Ronget, V., Ramirez-Loza, J.-P., Cubaynes, S., & Pavard, S. (2021). The kinship matrix: inferring the kinship structure of a population from its demography. *Ecol. Lett.*, *24*(12).

Goodman, L. A., Keyfitz, N., & Pullum, T. W. (1974). Family formation and the frequency of various kinship relationships. *Theor. Pop. Biol.*, *5*(1), 1–27.

Horn, A. R., & Johnson, C. R. (2010). *Matrix Analysis*. Cambrdige, CB2 8BS UK: Cambridge University Press.

Magnus, J. R., & Neudecker, H. (2019). *Matrix Differential Calculus with Applications in Statistics and Econometrics*. Oxford, OX4 2DQ UK: John Wiley & Sons, Inc.

Pullum, T. W. (1982). The eventual frequencies of kin in a stable population. *Demography*, *19*(4).

Roper, M., & Caswell, H. (2023). Kin in space [manuscript in preperation].

Tuljapurkar, S. D. (1982). Population dynamics in variable environments. iv. weak ergodicity in the lotka equation. *J. Math. Biol.*, *14*.