Time spent and left of transient states in stationary populations

Tim Riffe[*1], Francisco Villavicencio[2], and Nicolas Brouard[3]

[1]Max-Planck-Institute for Demographic Research
[2]University of Southern Denmark
[3]Institut National d'Études Démographiques

April 17, 2018

Abstract

Background   The Brouard-Carey equality establishes equality between the distributions of years-lived and years-left in stationary populations.

Objective   We aim to generalize this equality to account for time spent and time left in states within multistate stationary populations.

Results   We provide two intuitive proofs that the distribution of time spent and left in transient states is equal in multistate stationary populations.

Conclusions   This equality may be helpful under certain constraints to estimate the distribution of otherwise unobserved onset timing for health or other states.

1   Background

The Brouard-Carey equality establishes that the distributions of years-lived and years-left are identical in perfectly stationary populations (Brouard 1989, Vaupel 2009, Rao and Carey 2015, Villavicencio and Riffe 2016). Consider the case of multistate stationary populations, as implied by fixed transition rate schedules (over age) that govern mortality and the movement of individuals between a finite number of states. Multistate stationary populations are characterized by a fixed age-state structure (Perron 1907, Frobenius 1912), a property often exploited by Markov models with the goal of estimating state expectancies or similar. Since the age-state distribution is a fixed attribute, summing over states within age produces an age distribution that is also a fixed attribute, and it follows that the Brouard-Carey equality still holds in the aggregate with respect to age and remaining years of life distributions.

Individual discrete lifecourse trajectories resulting from fixed transition rate schedules consist in sequences of states. Such trajectories are not completely random processes, but these are constrained in the limit to abide by the laws of probability. Specifically, the probability of observing any particular discrete lifecourse trajectory, a Markov sampling path, is the product of the sequence of transition probabilities required to produce it,

---

[*]riffe@demogr.mpg.de

and the sum of the probabilities of all possible trajectories is one. The set of all possible trajectories (and their corresponding probabilities of occurring) — a potentially very large set of distinct trajectories — is also a fixed attribute of the stationary population.

Throughout this exposition we rely on the notion of a multistate stationary population; we therefore first offer a toy example of one to illustrate some of the properties that form the background to the symmetry property that we wish to describe. Define a two-state model of healthy (H) and sick (S) with three age classes, and including bidirectional flows between states as well as exits from each state to death (D). A set of example transition rates are given in Tab. 1. Each new birth also has a .9 (.1) probability of being born healthy (sick).

Table 1: Example transition rates in four discrete single age classes for transitions within and between good health (H) and sickness (S), and to death (D).

Age	HH	HS	HD	SS	SH	SD
0	0.89	0.10	0.01	0.10	0.70	0.20
1	0.70	0.20	0.10	0.20	0.50	0.30
2	0.50	0.30	0.20	0.30	0.30	0.40
3	0.00	0.00	1.00	0.00	0.00	1.00

Hold transition rate schedules fixed for a sufficiently long time, such that strong ergodicity takes hold, and hold the population to a constant size. Each new cohort is expected be identical, and the numbers of healthy and sick individuals in each age are expected to be the same in each time step. The set of potential discrete trajectories that might arise from these transition probabilities, and the probability of observing each, is given in Tab. 2:

Table 2: The set of all possible discrete transient trajectories and probability of observing each, given the transition rates in Tab. 1.

Trajectory	Probability	Trajectory	Probability
H	0.00900	SHHH	0.02450
S	0.02000	HSHH	0.02250
HH	0.08010	SSHH	0.00250
SH	0.00700	HHSH	0.04806
HS	0.02700	SHSH	0.00420
SS	0.00300	HSSH	0.00540
HHH	0.11214	SSSH	0.00060
SHH	0.00980	HHHS	0.16821
HSH	0.00900	SHHS	0.01470
SSH	0.00100	HSHS	0.01350
HHS	0.06408	SSHS	0.00150
SHS	0.00560	HHSS	0.04806
HSS	0.00720	SHSS	0.00420
SSS	0.00080	HSSS	0.00540
HHHH	0.28035	SSSS	0.00060
		Total	1.00000

For example, given a birth cohort of 100000 individuals, we would expect to see 900 individuals that are born healthy and then die before reaching the second age class. We

can force these proportions to be exact either via deterministic renewal of the population, or as a limiting property as the population size approaches infinity. More ages and states will make the number of potential trajectories factorially larger, but the sum of the probabilities of all trajectories will always be one. The important feature for the present is that the detailed discrete lifecourse composition of each entering cohort is under these conditions identical, uniquely defined by the transition rates, and therefore a feature of the multistate stationary population. By extension, the composition by state episodes (with respect to e.g. episode duration and timing) is identical under stationarity. This property forms the groundwork to the proof of our main result.

2   Relationship

We introduce an extension of the Brouard-Carey equality to states in multistate stationary populations. Formally:

Theorem 2.1. Given a multistate stationary population that has been stationary for a suffiently long time, the probability that an individual randomly selected at time t is in state s and has been in s continuously for at least x time is equal to the probability of being in state s and remaining in s for at least x time.

   Equivalently, the distribution of time spent in s is equal to the distribution of time left in s. This proposition might not be intuitive at first glance, so we prove it, and then speculate as to how this equality may be put to good use.

3   Two proofs of within-state symmetry

We provide two discrete proofs of theorem 2.1. Each may appeal to a differnt kind of intuition. The first is built up from the possible discrete life trajectories defined by a set of transition probabilities, such as those shown in Tab. 2. The second follows the Lexis aggregate approach of Villavicencio and Riffe (2016).

Proof 1 of Theorem 2.1. This proof follows five steps, which we illustrate along the way:

1. Complementarity identity: The lengths of two durations created by bisection of a single continuous duration sum to the length of the original duration. A duration can be represented as a line segment, potentially representing a state episode that comprises a subsection of a life-line. Points along a single within-person duration can be sampled over time in arbitrarily fine and regular time steps, δ. Each time step is a segment bisector, collecting two sets of values: 1) time spent in the state prior to the sample point, and 2) time left until exiting the state. If observations are evenly spaced, and values are truncated to the nearest δ, by way of complements these two sets will consist of the same values.

   If the duration of the jth episode of the ith potential life trajectory is called $d^{i,j}$, the age of entry is $a_L^{i,j}$ and the age of exit is $a_R^{i,j}$, such that $d^{i,j} = a_R^{i,j} - a_L^{i,j}$. The set of time-spent values that one would sample over the course of $d^{i,j}$, $A^{i,j}$ is defined as:

$$A^{i,j} = \left\{ \delta \cdot k \mid k \in \mathbb{Z}^+ , \ 0 \leq k \leq \left\lfloor \frac{d^{i,j}}{\delta} \right\rfloor \right\} \qquad , \tag{1}$$

where $\lfloor\ldots\rfloor$ denotes truncation to the lower integer bound. The set of time left values, T^i is also:

$$T^i = \left\{ \delta \cdot k \mid k \in \mathbb{Z}^+ , \ \left\lfloor \frac{d^{i,j}}{\delta} \right\rfloor \geq k \geq 0 \right\} \qquad . \qquad (2)$$

Figure 1 illustrates the construction of $A^{i,j}$ and $T^{i,j}$ in expressions (1) and (2). The central bisector moves along the duration from the time of entry until the time of exit, creating two sets of values.

Figure 1: A lifeline of the jth episode of a reference state individual i showing the construction of the sets $A^{i,j}$ and $T^{i,j}$.update superscripts to ^{i,j} in figure
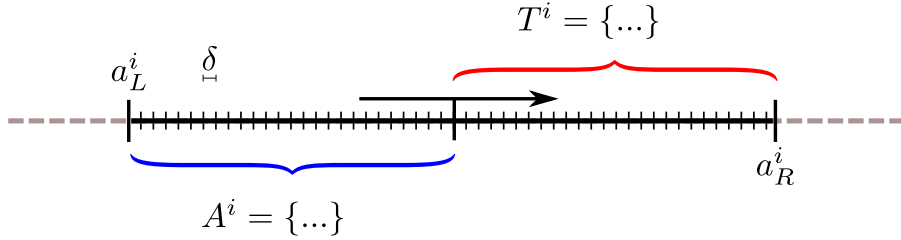
For example, from the 21st life trajectory in Tab. 2, HSSH, sickness might be the state of interest, in which case SS becomes the episode of interest in this trajectory. If each time step is a year, $\delta = 1$, then we have $A^{21,1} = \{0, 1, 2\}$ and $T^{21,1} = \{2, 1, 0\}$: Where order does not matter, these are two equal sets.

2. Complementarity identity over replicated life trajectories: Under stationarity, the probability of observing a given episode is replicated every δ time step, where δ is the same time step over time and age, i.e., the same spacing as our observation spacing in figure 1) but extended over a Lexis space. This can be represented with a set of age-aligned and identically long cohort segments placed side-by-side, spaced apart by δ. This potentially large set of segments may be imagined as a plane, although this would only hold in the limit. One could in this setting take a census at a single point in time, collecting a set of time spent and time left values ($A^{i,j}$ and $T^{i,j}$), each from a repeated episode in sequence. The two sets observed at a single time point but drawn from the replicating population of life trajectories will be identical to the first two sets that were sampled from a single duration over its entire duration, if the same δ-truncation is applied.

Figure 2 illustrates this notion with uniformly-spaced lifelines in a Lexis configuration. The vertical line indicates a hypothetical census at time t of the population of this replicated life trajectory. At time t, the blue-highlighted segments indicate the set of time-spent values in $A^{i,j}$, and red-highlighted segments are the time-left elements of $T^{i,j}$.

Formally, sets $A^{i,j}$ and $T^{i,j}$ consist of the same values as the previous, but coming from distinct instances of life trajectories generated in a uniform series from between $t-a_R^{i,j}$ years ago until as recently as $t-a_L^{i,j}$ years ago. This demonstrates both period-cohort set equality and time spent-left set equality. The blue and red triangles in Figure 2 are simple rotations of one another.

Figure 2: The jth episode of the ith life trajectory is repeated in δ time steps. A census with followup now constructs the sets $A^{i,j}$ and $T^{i,j}$ with values identical to the within-individual sets.
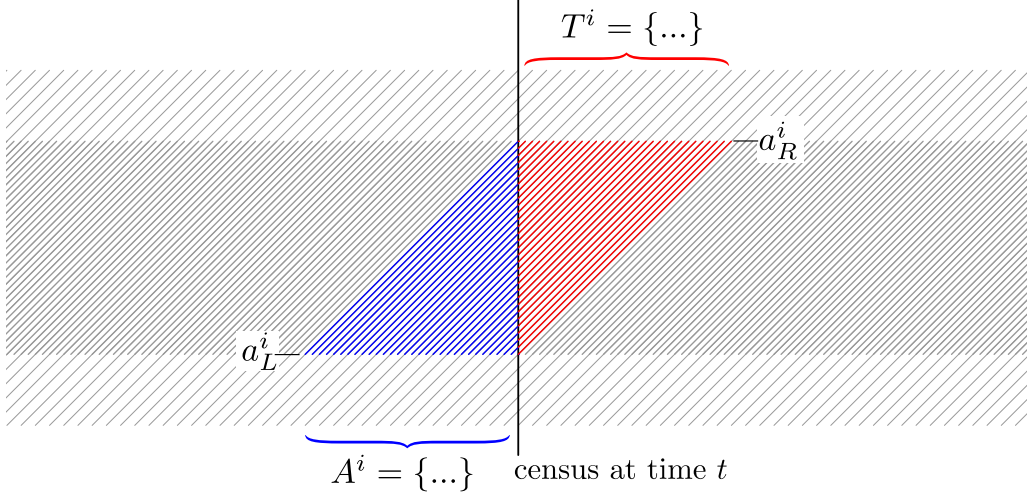
Continuing with the episode $d^{21,1}$ in Tab. 2, if we replicate this trajectory in single year steps, and take a census in year t, the values of $A^{21,1}$ and $T^{21,1}$ are the same as before $\{0, 1, 2\}$, but taken from three different life trajectories that began in years $t - 4$, $t - 3$, and $t - 2$, and were observed at ages 3, 2, and 1, respectively.

3. Identical set concatenation identity: Repeat the construction of time spent and time left values for some other reference episode, $d^{i',j'}$, as in step 1. This episode might have a different duration, their sets of time-spent, $A^{i',j'}$, and time left $T^{i',j'}$ will be equal to one another but potentially distinct from those drawn from the previous trajectory-episode. $A^{i,j}$ and $T^{i,j}$ range from 0 to $d^{i,j}$, whereas $A^{i',j'}$ and $T^{i',j'}$ range from 0 to $d^{i',j'}$. However, their concatenations are identical:

$$\{A^{i,j}, A^{i',j'}\} = \{T^{i,j}, T^{i',j'}\} \tag{3}$$

The concatenation identity also applies if we replicate the life trajectory $d^{i',j'}$ in δ time steps as in step 2. It is also true if we weight the sets derived from each trajectory by the probability of observing each trajectory.

For example, the 22nd trajectory from Tab. 2, SSSH, has a single episode of sickness of duration three, yielding identical sets $A^{22,1} = \{0, 1, 2, 3\}$ and $T^{22,1} = \{3, 2, 1, 0\}$. If the trajectory is replicated as in step 2, the observations of $d^{22,1}$ will come from four life trajectories born in years $t - 5$ through $t - 1$ and captured at ages 3, 2, 1, and 0, respectively. Whether observed over time as in step 1 or over individuals within time, as in step 2, the concatenation of $A^{21,1}$ and $A^{22,1}$ is equal to the concatenation of $T^{21,1}$ and $T^{22,1}$, $\{0, 0, 1, 1, 2, 2, 3\}$. If each birth cohort derived from the probabilities in Tab. 1 starts with 100000 individuals, we would expect a hypothetical census to yield 540 instances of trajectory 21 and 60 of 22, so we would expect 600 observations of values 0, 1, and 2, and 60 observations of 3.

4. Distribution equality by induction By induction we can keep adding state episodes from the stationary population, constructing A and T for each, and concatenating

over episodes as in step 3. The concatenation of all episode-specific time-spent sets and the concatenation of the corresponding time-left sets will always be identical. This is so both within trajectories and over time and over trajectories observed at a single point in time. Therefore the probability of selecting a particular value from the concatenated time-spent set is identical to the probability of selecting the same value from the cocnatenated time-left set.

$$\square$$

Our point of departure was that of a perfectly stationary population with identically distributed lifelines that aggregate to a fixed age-state structure, and this is also the result that we arrive at by induction in the final step of the proof. In this strict environment it is therefore the case that our theorem holds. A more rigorous proof would allow δ to decrease to 0 in the limit. The present approach does not allow this because probability statements would no longer hold. We may require a population of infinite size in order to make the jump from a fixed age-state structure to an invariant set of lifelines, where each lifeline is itself an invariant sequence of states.

4   Discussion

The approach from this proof is equally valid to prove the original statement of the Brouard-Carey equality (where the state is "alive"), but it is more general. The statement and proof is flexible enough to hold for irreversible and reversible states. It also applies to repeatable states, whether time spent in the state is kept in cumulative fashion over spells, or whether the clock resets to zero on each entry into the state. The equality may also be conditionable in curious ways: for example, the distribution of time-spent in a state conditional on having entered at age a must also be equal to the distribution of time-left in the state, conditional on having entered at age a. Likewise, one may condition statements on exit age. One may also arbitrarily merge states, and the equality still holds within the newly merged state.

At first glance, this equality is probably less intuitive than the original Brouard-Carey equality, because state entry is not necessarily aligned on age zero. It is less visible in commonly-produced plots because plots of stacked sequences look chaotic, often even if clustered or sorted. It might be tempting to think that due to state-varying vital rates, the equality simply ought not hold. However, the basis of the proof is the observation that if each individual duration is symmetrical by complements, then so are aggregations of durations, irrespective of alignment. Since each cohort in a perfectly stationary population of infinite size is an identical copy of the previous, census-like cross-sections are also equally-composed.

5   Potential applications

Empirical applications of the presently-described transient tenure equality may be easy to conjure up. For example, imagine a hypothetical health state that shows no noticeable symptoms, but that is medically measurable. One may take a census with regular follow-ups, until eventually the state is exited by each individual, whether by absorption into death or entry into another state. Then, if the assumption of stationarity is acceptable, one may be able to say something about onset timing in the aggregate, itself unobserved.

6 Simulation

[section commented out until exercise more completely designed]

7 Conclusion

Stationary populations are more symmetrical with respect to time lived and spent than has been previously described. We have given an intuitive proof that the distribution of time spent and left is equal within states in stationary populations. We do not presume that this relationship will be immediately able to answer pressing questions, but we hope that it may inspire new approaches in empirical measurements. We think that that researchers working with left-censored or truncated data or any kind of multistate models should be generally aware of this equality in case it may come in handy as a heuristic.

The relationship between years lived and left established by the Brouard-Carey equality is itself amenable to non-zero growth rates (Riffe 2015). We speculate that the transient tenure equality we describe here may also be tractable in stable populations, but we have not investigated this possibility in detail. Future research, potentially more sophisticated simulations, should establish the impact of departures from stationarity and establish the fuzzy bounds of usability for this relationship.

References

Brouard N (1989) Mouvements et modèles de population. Institut de formation et de recherche démographiques

Frobenius GF (1912) Über Matrizen aus nicht negativen Elementen. Königliche Akademie der Wissenschaften

Perron O (1907) Zur theorie der matrices. Mathematische Annalen 64(2):248–263

Rao ASS, Carey JR (2015) Generalization of Carey's equality and a theorem on stationary population. Journal of Mathematical Biology 71(3):583–594

Riffe T (2015) Renewal and stability in populations structured by remaining years of life. MPIDR Working Papers (WP-2015-007)

Vaupel JW (2009) Life lived and left: Carey's equality. Demographic Research 20:7–10, 10.4054/DemRes.2009.20.3

Villavicencio F, Riffe T (2016) Symmetries between life lived and left in finite stationary populations. Demographic Research 35(14):381–398