RESEARCH

Prevalence and the variance of state occupancy time

Tim Riffe*

This manuscript is in its early stages

Correspondence:
riffe@demogr.mpg.de
Max-Planck-Institute for Demographic Research, Konrad-Zuse-Str. 1, 18057 Rostock, Germany
Full list of author information is available at the end of the article
*

Abstract

Background: Markov reward methods have been proposed to calculate the variance of state occupancy time based on age-structured prevalence and survivorship.

Objectives: I aim to clarify the assumptions of this approach, give bounds to its reasonableness, and suggest improvements.

Methods: I calculate results for extreme cases to provide bounds, I simulate the variance under simple assumptions, and simulate a more natural inter-individual state distribution.

Results: I show that state occupancy variance for Sullivan-style inputs is not identified, and I show where previously proposed methods fall with respect to reasonable bounds, randomly generated variances, and my own opinion about what a reasonable inter-individual state distribution might look like.

Conclusions: The variance of state occupancy time is only identified if a) state life trajectories are directly observed or b) a process model, such as an incidence-based Markov model, is specified. Sullivan-calculations of life expectancy do not imply a single variance, and are therefore insufficient to make statements on inter-individual disparities in state occupancies.

Keywords: Healthy life expectancy; Sullivan method; Healthy life variance

Introduction

Healthy inquality is usually measured between populations by comparing life expectancies, health expectancies, or poor health expectancies. Within populations there is also inquality in health outcomes, either due to violations of the assumption of homogeneity of risk sets or due to random variation between people otherwise subject to the same risk. If there

bla bla I'll end up proposing some alternative variance calcs for Sullivan situations, but also that these are always hypothetical and that in fact the variance is not determined, and has a rather wide potential range of values depending on actual life trajectories.

Sullivan in continuous time

Let's begin with some familiar notation, $\ell(x)$ denotes lifetable survivorship, with an arbitrary radix, or starting population. If $\ell(0) = 1$ then $\ell(x)$ can be interpreted as a probability of surviving from birth until age x. $\pi(x)$ is the prevalence of a given condition (healthiness or unhealthiness, disability) at age x, and it falls in the range

$[0,1]$, giving it a probability interpretation. The commonly used Sullivan method [1] of calculating healthy life expectancy, $e^H(x)$ is to integrate the product of these two functions.

$$e^H(x) = \frac{1}{\ell(x)} \int_x^{\omega} \ell(t)\pi(t)\mathrm{d}t \tag{1}$$

Replace π with its complement to arrive at the complementary expectancy, $e^U(x)$, such that overall life expectancy for this age is $e(x) = e^U(x) + e^H(x)$, a rudimentary composition.

Two assumptions

If we would like to know something about the variance of state occupancy in a Sullivan setup, then we must make some assumptions about how the state is distributed over individuals. The Sullivan method does not state which between-individual state distribution underlies prevalence in a given age. If $\pi(x) = 0.5$, shall we assign each individual age x a value of 0.5 a so-called *fixed reward*, or shall we assign half of them a value of 1, and the other half a value of 0, a *Bernoulli reward* [2]? Or something else entirely? There are infinitely many ways to distribute state π among individuals in age x such that the value $\pi(x)$ is maintained. By extension, the inter-individual distribution of total time spent in state π is also infinitely variable, which implies that the variance and other moments of implied state occupancy are not uniquely identified.

Expressing the lifetable as a Markov chain with *rewards* defined as prevalence gives a unique definition for the variance of state occupancy for each of the above assumptions, but it is unclear to me how well-supported these are. Many other reward types are discussed by Caswell & Zarulli [2], and I do not discuss these, nor do I treat the case of multistate populations. Even so, the conclusions reached here will generalize to the case of multistate populations.

Bernoulli rewards

I'll first give my own lifetable deconstruction of the matrix algebra approach given to Bernoulli state variance in [2]. Then $e^H(x)$ is defined per (1), and it is the first moment of state occupancy, which we can denote $\eta^{(1)}$, where the superscript in parentheses denotes the moment number and is not a power. We continue to calculate the second moment of state occupancy, $\eta^{(2)}$ as:

$$\eta_x^{(2)} = \frac{1}{\ell_x} \sum_{t=x}^{\omega} \ell_t \left[\pi_t + 2(1-q_t)\pi_t\eta_t^{(1)} \right] \quad , \tag{2}$$

where q_x is the probability of death in the interval and ω is the highest age of death. Here notation omits intervals, but we assume to be working in discrete bins of uniform width, where death and transition changes only happen in the moment of interval steps, implying a stepped survival curve and discrete life trajectories. We also assume that the same prevalence applies to each length-of-life bin within each age, ergo that all other unmodelled population strata have the same prevalence.

Then the variance is determined, and it can be calculated as:

$$Var_x = \eta_x^{(2)} - (\eta_x^{(1)})^2 \tag{3}$$

I highlight two assumptions behind this expression: 1) mortality is independent of whether one is in the prevalent state, and 2) the same Bernoulli prevalence extends to any further stratification of the members in a given age class, x.

Fixed rewards

If instead of assuming that prevalence is partioned in a binary fashion, and each individual experiences π_x fraction of the year in the state, then equation (2) becomes

$$\eta_x^{(2)} = \frac{1}{\ell_x} \sum_{t=x}^{\omega} \ell_t \left[\pi_t^2 + 2(1 - q_t)\pi_t \eta_t^{(1)} \right] \quad , \tag{4}$$

and (3) is the same. State occupancy variance under the assumption of fixed rewards also has a more intuitive lifetable expression:

$$Var_x = \frac{1}{l_x} \sum_{x}^{\omega} (\mathcal{P}_t - \eta_x^{(1)})^2 d_t \quad , \tag{5}$$

where $\mathcal{P}_t$ is the cumulative prevalence from age x to age t, and d_t is the probability of survival from age x to age t, defined as $(q_t l_t)/l_x$.

0.1 Illustrations of Bernoulli and fixed prevalence

Equations (**??**)

Competing interests
The authors declare that they have no competing interests.

Author's contributions
TR did everything.

Acknowledgements
Thanks to Douglas Wolf and Jennifer Karas Montez for posing a question that led to this work, and to Hal Caswell and Virginia Zarulli for the groundwork on which this is based and for fielding inquiries.

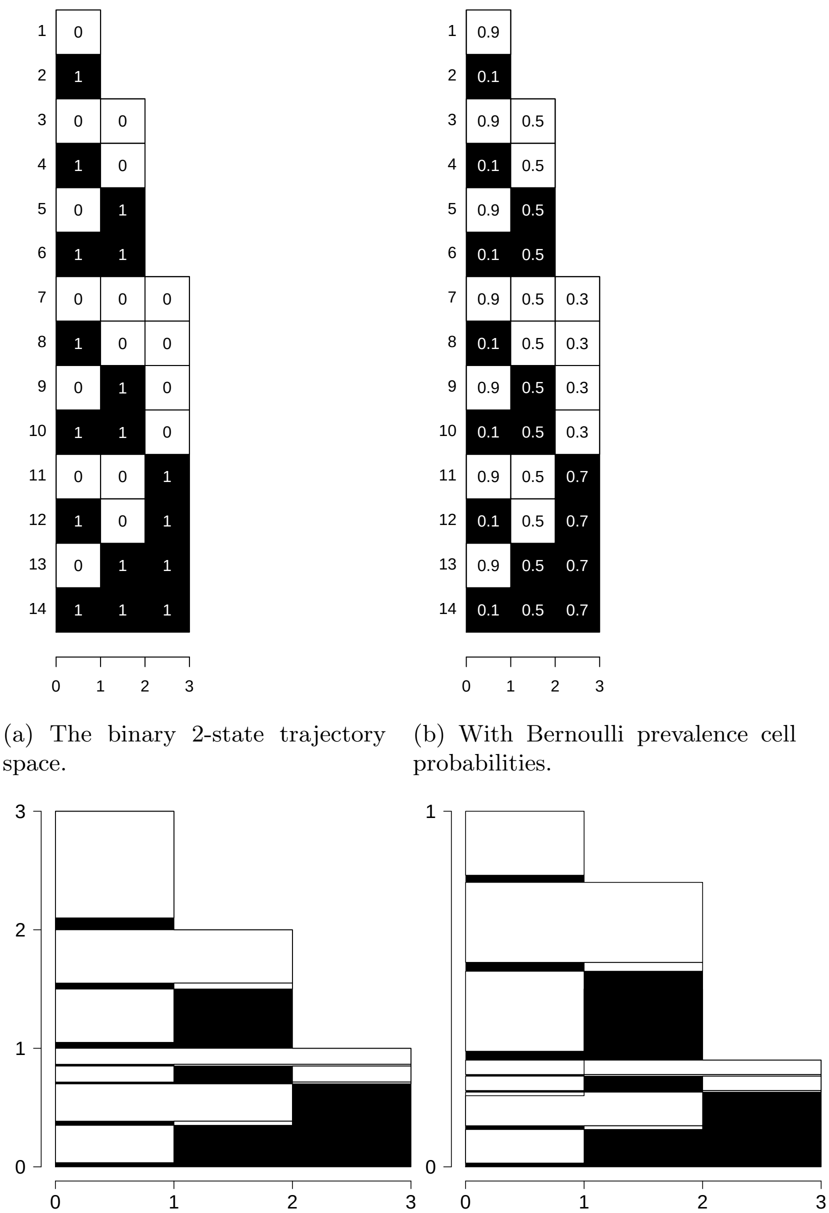
References

1. Sullivan, D.F.: A single index of mortality and morbidity. HSMHA health reports **86**(4), 347 (1971)
2. Caswell, H., Zarulli, V.: Matrix methods in health demography: a new approach to the stochastic analysis of healthy longevity and dalys. Population health metrics **16**(1), 8 (2018)

Additional Files

Additional file 1 — Sample additional file title
Additional file descriptions text (including details of how to view the file, if it is in a non-standard format or the file extension). This might refer to a multi-page table or a figure.

Additional file 2 — Sample additional file title
Additional file descriptions text.



(a) The binary 2-state trajectory space.

(b) With Bernoulli prevalence cell probabilities.

(c) With trajectories Bernoulli weighted, conditional on length of life.

(d) With trajectories Bernoulli and lifetable weighted.

Figure 1: A depiction of the Bernoulli life trajectories implied by the prevalence and lifetable of Tab. ??