Exact Poisson confidence intervals for life expectancy*

Michelle Deville[†] Tim Riffe[‡] Andrew Noymer[†,§]

25 September 2015

Abstract

Life expectancy statistics are almost never presented as interval estimates (i.e., with confidence bounds). For national-level populations, life expectancy is very sharply estimated. Except for the smallest countries, population size and deaths are large enough that the confidence intervals are so narrow as to be unhelpful, and are therefore omitted. However, in very small countries and in sub-national polities, deaths may be few enough that confidence intervals for life expectancy can provide informative uncertainty estimates. We present a method for calculating exact Poisson-based confidence intervals, based on theoretical work by Brillinger (1986). We illustrate this method using data from three real polities: Guam, American Samoa, and Saipan. For these populations, the exact CI is meaningfully different from — and more conservative than (i.e., wider) — the standard formula by Chiang (1984). We also analyze simulated populations to calculate response curves, illustrating which demographic conditions justify exact methods.

---

*PAA 2016 extended abstract. Please contact authors for latest version.

[†]Program in Public Health, University of California, Irvine.
[‡]Max Planck Institute for Demographic Research, Rostock, Germany.
[§]Correspondence to: noymer@uci.edu

Introduction

Life expectancy statistics are almost never presented as interval estimates (i.e., with confidence bounds). For national-level populations, life expectancy is very sharply estimated. Except for the smallest countries, population size and deaths are large enough that the confidence intervals are so narrow as to be unhelpful, and are therefore omitted. However, in very small countries and in sub-national polities, deaths may be few enough that confidence intervals for life expectancy can provide informative uncertainty estimates. We present a method for calculating exact Poisson-based confidence intervals, based on theoretical work by Brillinger (1986). We illustrate this method using data from three real polities: Guam, American Samoa, and Saipan. For these populations, the exact CI is meaningfully different from — and more conservative than (i.e., wider) — the standard formula by Chiang (1984). We also analyze simulated populations to calculate response curves, illustrating which demographic conditions justify exact methods.

Data and methods

The core of this paper is to calculate interval estimates of life expectancy two ways, and to compare them. The first way is the standard method, based on large-sample asymptotics, described in enormous detail by Chiang (1984). The second way, described immediately below, is extremely straightforward but is computationally expensive. Modern computers can calculate the second method in perfectly reasonable amounts of time (e.g., all of table 1 took under 40 seconds to compute, on a laptop from 2012 and with 100,000 replications per life table). However, it is not hard to see why this method was not preferred at the time when Chiang (1984) was written, since the fastest computers typically available in academic computing environments at that time were machines like the Digital VAX 11/750, much much slower than computers nowadays, even laptops.

Our method

Statistical theory by Brillinger (1986) posits that death counts follow a Poisson distribution. Our method uses this to build a simple way of calculating a CI for life expectancy. We take the vector of age-specific ${}_nD_x$ values (death counts), and we replace each element of that vector with a random Poisson deviate with mean value ${}_nD_x$ at that age (x) value. Once we have a complete ${}_nD_x$ vector from that method, we calculate life expectancy in the usual way (i.e., computing a life table), taking the exposure values ${}_nK_x$ as fixed. We then repeat the whole process a reasonable number of times (100,000, herein), and get, therefore 100,000 $e(0)$ values. Our interval estimate is then bounded by the 2.5 and 97.5 percentiles of the resulting set of $e(0)$ values. Our approach is, clearly, very computationally intensive (table 1 represents 1.2 million life tables being constructed). But, as noted above, modern computers are easily up to the task. What is more, our approach is, algorithmically, very simple: this paragraph is, for all intents and purposes, a complete description of the technique.

Since it is not possible to specify a Poisson deviate with mean zero, if there are no deaths in an age group, we use a Poisson deviate with a fractional mean. An alternate approach would be always to use 0 in such cases; this may be logical from a purely statistical point of view, since zero mean implies zero variance, but it seems to make more sense demographically if some nonzero variance is permitted at all ages.

Comparisons

I) Simulations

We will simulate populations of various sizes to see at what point, under various conditions, the difference between Poisson and the Chiang approach becomes negligible. We will vary the conditions by the following parameters:

1. Population size
2. Population age-structure
3. Mortality pattern (Coale-Demeny N, S, E, W)
4. Mortality level

II) Real populations

We have calculated our comparisons for three real polities: the Pacific Territories of the United States. These are Guam, American Samoa, and Saipan (or CNMI, the Commonwealth of the Northern Mariana Islands). For each population, males and females were calculated separately, and two time periods were used, 2000 and 2010. The death counts are published by NCHS and the Census is the source of the ${}_nK_x$ values. Following the practice of NCHS decennial life tables, we used 3 years of deaths, centered on the census year (e.g., 1999–2001 for 2000); the exposure values are then adjusted accordingly so that a correct life table is produced. This gives more deaths and a more reliable mortality estimate for small populations.

The advantage of using these populations is that they are small enough potentially to have meaningful differences between the CI methods; US states are, in general, far too large (the CIs, for either method, become extremely narrow). Moreover, counties, which are about the right size, do not have publicly-available death counts by age and sex (nor, even, states after 2004).

Results & Discussion

The part of the analysis on the simulation study is not complete as of September 2015.

The part of the analysis on the real populations is complete and is presented below. While this component of our study accounts for less variation than can be accommodated by the simulation study, the empirical component below is, arguably, more important because real populations are where the rubber meets the road, so to speak.

FEMALES	2000			2010		
American Samoa	LB	CE	UB	LB	CE	UB
Chiang	73.58	74.32	75.07	72.69	73.34	73.99
Poisson	73.28	74.34	75.40	72.34	73.36	74.38
population; deaths	*28,027*	*286*		*27,349*	*343*	
CNMI	LB	CE	UB	LB	CE	UB
Chiang	73.60	74.43	75.25	75.73	76.47	77.21
Poisson	73.21	74.45	75.75	75.45	76.49	77.55
population; deaths	*37,237*	*197*		*26,137*	*192*	
Guam	LB	CE	UB	LB	CE	UB
Chiang	75.78	76.18	76.59	75.67	76.10	76.53
Poisson	75.63	76.19	76.74	75.55	76.10	76.65
population; deaths	*75,624*	*785*		*77,790*	*983*	
MALES	**2000**			**2010**		
American Samoa	LB	CE	UB	LB	CE	UB
Chiang	68.41	69.08	69.74	68.78	69.44	70.11
Poisson	68.03	69.09	70.18	68.45	69.45	70.49
population; deaths	*29,264*	*417*		*28,170*	*467*	
CNMI	LB	CE	UB	LB	CE	UB
Chiang	71.93	72.76	73.58	72.36	73.09	73.82
Poisson	71.63	72.78	73.99	72.05	73.10	74.18
population; deaths	*31,984*	*248*		*27,746*	*334*	
Guam	LB	CE	UB	LB	CE	UB
Chiang	70.40	70.86	71.31	70.01	70.46	70.90
Poisson	70.24	70.86	71.49	69.85	70.46	71.07
population; deaths	*79,181*	*1,242*		*81,568*	*1,549*	

Table 1: Life expectancy estimates and confidence intervals, US Pacific Territories, 2000 & 2010. LB=lower bound; CE=central estimate; UB=upper bound. Population denotes census population. Deaths are for three years, centered on the census year; when calculating life tables, exposure has been adjusted accordingly. If calculated directly, the population and death counts in the table will give a crude death rate $\approx 3\times$ too high.

In these real populations, we observe interval estimates that are modestly, but meaningfully, wider, using the Poisson method than with Chiang's formula. For example, for Saipan (CNMI) females for 2000, the Poisson CI is 0.89 year wider than the CI calculated with Chiang's method. This is less than a year, but given that the Chiang CI is only 1.65 year wide, the Poisson CI is over half-again as wide. That is clearly a non-trivial difference.

Conclusion

The standard method of Chiang performs poorly for small populations. It produces interval estimates that are too narrow relative to Poisson simulation. Statistical theory favors the Poisson approach over the asymptotic approach of Chiang, so we are confident that our interpretation of the results — that Chiang's intervals are too narrow, as opposed to Poisson being too wide — is correct.

Works Cited

Brillinger, David R. 1986. "The natural variability of vital rates and associated statistics. With discussion." *Biometrics* 42(4):693–734.

Chiang, Chin Long. 1984. *The life table and its applications*. Robert E. Krieger Publishing, Malabar, Florida.