CSE383M and CS395T Problem Set

due Wednesday, February 22, 2012

Note: In the spirit of everything in this course, these problems may not be quite "well-posed". You're supposed to make any reasonable assumptions that you need to make, as if this was real data and you really need to get an answer.

Actuarial population mortality is well summarized as follows: For people of age t (in years), the probability of death in the one year period between t and $t+1$, the so-called "hazard function" or "death rate" $\lambda(t)$, is

$$\lambda(t) \approx \frac{1}{1+e^{a-bt}} \tag{1}$$

where

$$a \approx 11.5, \qquad b \approx 0.11 \text{ (per year)} \tag{2}$$

1. Answer the following questions by *simulation* of a population of individuals.

(a) What is the median life expectancy at birth?

(b) What is the mean life expectancy at birth?

(c) Plot a histogram of life expectancy at birth.

(d) If Country A has a constant number of births per year, what is the median age of its population?

(e) If Country B has population that is increasing at the rate of 5% per year, what is the median age of its population?

2. Answer these questions analytically in terms of a and b. (Hint: I used Mathematica. If you find the same Mathematica bug that I did, you'll need to recognize that, for positive x,

$$\log(-x) = \log(x) + i\pi \tag{3}$$

and that the imaginary piece might just be a removable constant of integration. Also recall that you can legally download Mathematica for free from the UT Austin Physics Department – link is on the course wiki.)

(a) What is the "survival function" $S(t)$, defined as the fraction of people born who are still alive at age t (in years), in terms of a and b?

(b) What is the analytical relationship between the survival function $S(t)$ and the distribution of life expectancy at birth $p(t)$, defined as the probability (at birth) of dying at age t?

(c) What is $p(t)$ analytically in terms of a and b?

(d) For the numerical values of a and b given above, compute the mean and median life expectancies to 8 significant figures, either analytically or numerically. (The requested accuracy is only to force you NOT to do this by simulation!)

3. A small town in Italy, with a stable population of 1000, is famous for its long lifespans. (They say it is from drinking a gallon a day each of red wine.) Its two oldest residents claim to be 102 and 106 years old, respectively.

(a) As a frequentist, can you rule out the null hypothesis that the town's mortality rate is actually the same as everyone else (i.e., equations 1 and 2)? With what p-value? (All methods are fair game: analytic, numerical, sampling, simulation.)

(b) As a Bayesian, consider the hypotheses that the town's value of b is different from 0.11. (This would be equivalent to re-scaling everyone's age, that is, slowing down aging by a constant factor.) What is the Bayesian probability distribution of b if the claimed residents' ages are correct. (Here also, all methods are fair game.)

(Hint: This problem isn't so easy, because you have to figure out the implications that the ages specified are the two *oldest* out of 1000 – a so-called "order statistic". Try for even a crude approximation if you can't figure out how to do it accurately. You might also think about how to random-sample the ages of the oldest two in a population of 1000 without having to simulate all 1000 people!)