B.6 Modern Survival Problem sheet 6: Censoring and truncation, frailty and repeated events

To be turned in by noon on 15 January, 2016

(1) A sample of patients taking a new blood pressure medication is asked whether they have experienced any vertigo since they started taking it; and if so, when the symptoms were first noticed. Some have not experienced symptoms yet, some report an exact time (in weeks after starting treatment), and some only say they know it was before a certain time.

Table B.1: Reports of vertigo

weeks	# taking it this many weeks who never had symptoms	# whose symptoms started at this many weeks	# whose symptoms started before this many weeks
1	45	6	0
2	22	11	0
3	23	10	0
4	19	22	3
5	12	37	2
6	10	33	6
7	3	16	9
8	5	13	4
9	3	8	9
10	0	9	15

Which observations are left-censored? Right-censored? Estimate the survival function (that is, probability of remaining symptom-free for x weeks)

(a) Ignoring the left-censored observations;
(b) Ignoring the right-censored observations;
(c) Taking all observations into account.

(2) In order to control the spread of a virus in a wild population, researchers spread food items laced with a vaccine. Once a week they capture a small number of animals and test whether they have developed an immune response

week	1	2	3	4	5	6	7	8	9	10
number sampled	5	4	7	3	4	6	3	8	5	4
number immune	0	1	2	0	2	1	2	4	4	3

Estimate the probability of being immune at week t

(a) using an exponential model;

(b) using a Weibull model;

(c) using the nonparametric MLE.

(3) A population has multiplicative frailty, so that the mortality rate for individual i is $B_i\alpha(x)$ at age x, where the B_i are i.i.d. positive random variables and $\lim_{x\to\infty}\alpha(x) = \infty$.

(a) Show that the population mortality goes to ∞ as $t \to \infty$ if the distribution of B_i is bounded away from 0.

(b) Show that the population mortality converges to a finite constant as $t \to \infty$ if the distribution of B_i has nonzero density at 0 and the hazard rate does not grow too quickly as $x \to \infty$. Give a formal condition for what "too quickly" would be.

(c) Suppose now that the baseline hazard is Gompertz, i.e., $\alpha(x) = e^{\theta x}$.

i. If the B_i have Gamma distribution with parameters (r, λ) — λ is the rate parameter — compute the population mortality rate $\mu(t)$ at age t.

ii. What is the hazard ratio between a subpopulation whose frailty has Gamma distribution with parameters (r, λ) and one with parameters (r', λ)?

(4) The paper [ZKJ07] includes a dataset, available to download from the Journal of Statistical Science, on the healthcare demand of 4406 patients in the public old-age health insurance scheme Medicare in the US. When you load this file in, the data will be in a data-frame `DebTrivedi`.

(a) The number of physician office visits is enumerated in the variable `ofp`, while `numchron` gives the number of chronic conditions, and `health` gives self-reported health (poor, average, excellent). Do one or more exploratory plots to illustrate the distributions of these variables, and their relationship.

(b) Fit a Poisson regression model to predict the number of office visits as a function of `health`, `numchron`, `gender`, `school` (number of years of schooling), and `privins` (indicator of whether the patient has private insurance). Interpret the result.

(c) Explain why you might want to fit a negative binomial model instead. Do the fit, and interpret the result.