

**BIOSTAT/STAT 576**  
**Statistical Methods for Survival Data**

**Problem Set 8**

1. Please refer to the manuscript by Chen (2009) appearing in *Biometrics*. In this manuscript, Chen (2009) deals with size-biased outcomes when the outcomes are intercept-sampled without censoring. Let's consider the following situation when censoring is involved in an intercept-sampling.

Suppose that  $T = W + S$  is the time-to-event between disease onset and death. Here,  $W$  is the time measured between the disease onset and the intercept-sampling observation time, while  $S$  is the time between the intercept-sampling observation time and the death. Denote that  $C$  is the censoring time of  $S$  measured from intercept-sampling observation time.

For a particular study, suppose that we would collect the data of  $(W_i, X_i, \Delta_i, Z_i)$  for  $n$  subjects, where  $X_i = \min(T_i, C_i)$ ,  $\Delta_i = I(T_i \leq C_i)$  and  $Z_i$  are the univariate time-independent covariates, such that

$$\log T_i = -\beta Z_i + \epsilon_i.$$

Here,  $\epsilon_i$  follow the same distribution of  $F(\cdot)$ .

- (a) Discuss potential complications in analysis of this type of data.
- (b) Propose some statistical methods to analyze the data. You are not limited to one proposal.
- (c) For the proposed method of your preference, discuss
  - what kind of assumptions would you need?
  - how are you going to justify your method?
  - what is your simulation plan?
  - what are the pros and cons of your method?
  - is there any room for improvement?