Survival Analysis with Accelerated Failure Time

- Models time to an event of interest
- The label is always positive
- Label may not be fully known, and could be censored
- For example, the time to death may be a range instead of an exact number
  - Sensored
- 4 kinds of sensoring
  - Uncensored
  - Right censored, [a, infinity] where a is the lower bound
  - Left-censored, label is in form [0,b] where b is the upper bound
  - Interval censored [a,b]
- Right censoring is most common

Accelerated Failure Time Model

Accelerated Failure Time (AFT) model is one of the most commonly used models in survival analysis. The model is of the following form:

$$\ln Y = \langle \mathbf{w}, \mathbf{x} \rangle + \sigma Z$$

where

- $\mathbf{x}$ is a vector in $\mathbb{R}^d$ representing the features.
- $\mathbf{w}$ is a vector consisting of d coefficients, each corresponding to a feature.
- $\langle \cdot, \cdot \rangle$ is the usual dot product in $\mathbb{R}^d$.
- $\ln(\cdot)$ is the natural logarithm.
- Y and Z are random variables.
  - Y is the output label.
  - Z is a random variable of a known probability distribution. Common choices are the normal distribution, the logistic distribution, and the extreme distribution. Intuitively, Z represents the "noise" that pulls the prediction $\langle \mathbf{w}, \mathbf{x} \rangle$ away from the true log label $\ln Y$.
- σ is a parameter that scales the size of Z.