$0 < t_1 < t_2 < \cdots < t_r$: distinct observed death times 12/5/22

n_i : the size of the risk set at time t_i i.e., the number of individuals that were alive ~~bef~~ just before time t_i.

for $i < r$, $n_{i+1} = n_i - d_i - c_i$

d_i: # who died at time $t = t_i$

c_i: # number censored at times in $[t_i, t_{i+1})$.

Kaplan-Meier estimate: $\hat{S}(t) = \prod_{t_i \le t} \left(1 - \frac{d_i}{n_i}\right)$

Greenwoods confidence interval (1926)

$$\hat{S}(t) \pm z_{\alpha/2} \sqrt{\widehat{Var}[\hat{S}(t)]}, \quad \widehat{Var}[\hat{S}(t)] = \hat{S}(t)^2 \sum_{t_i \le t} \frac{d_i}{n_i(n_i - d_i)}$$

~~$n_i \to d_i$~~

First find $\lim_{d_i \to \infty}$ of $\hat{V}_{d_i}$

$$0 \le \hat{V} \le \sum_{t_i \le t} \frac{d_i}{n_i(n_i - d_i)}$$

We should assume that ~~each~~ as the samples grow d_i approaches some proportion of $n_i \to \infty$ say α_i, so

Very loose here

$$\sum_{t_i \le t} \lim_{n \to \infty} \frac{\alpha_i n_i}{n_i(n_i - \alpha_i n_i)} = 0, \text{ since it's} \sim \frac{n_i}{n_i^2}$$

The survival function can be expressed in terms of the stationary distribution. $=\{\pi_x\}_{x\in X_n}$

12/5/22

Let $\pi = [\pi_{\bar{0}}, \dots, \pi_{\bar{I}}]$ be the stationary distribution of the modified chain. For any $i \in [n]$, let m be a message injected at node i and T be m's extinction time (any value in $\mathbb{N}$).

~~If T=1, then~~

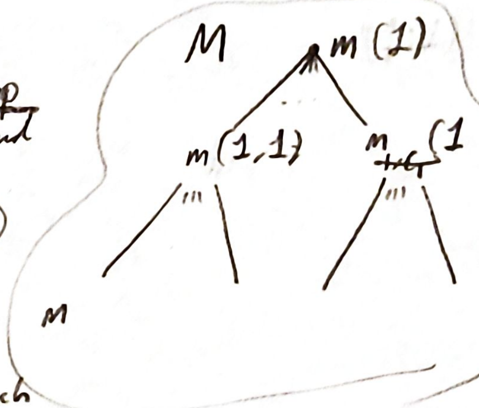
I also need a better notation for this set cause then Σ notation is usable

$\mathbb{P}(T=1) = \mathbb{P}(\text{a node in } \overleftarrow{N}_i \text{ was active}) = \text{sum}\{\pi_x : \boxed{\exists h \in \overleftarrow{N}_i \text{ s.t. } x_h = 1}\}$

$\mathbb{P}(T=2) = $ ~~$\mathbb{P}(T \neq 1)\mathbb{P}$(every child of m dies in a collision with another child or injection)~~ $(1-p_1)$

~~$= (1-p_1)[\mathbb{P}$(every child collides with another child) $+ \mathbb{P}$~~

→ b/w message survival time and graph parameters



Paper format.

Abstract

- The purpose of this paper is to see the relationship
- Say what we can about the theoretical answer.
- Take a statistical approach to understand the relationship

I need a way of indexing this tree that makes it easier to write out the probabilities.

Intro — Modified chain here.
Open form expression of the stationary distribution π
Expression of the survival function in terms of π.
Kaplan-Meier justification
Graph feature selection
Regression results → If positive (pray)
Discussion

Kaplan-Meier non-parametric estimate → Parsimonious parametric estimate → Regress parameters on graph features

Some error ε_1 | Some error ε_2 | Some error ε_3