$$N(x_i) \mapsto \begin{cases} rTS(x_i)U(x_i) & \text{if } x_i < \tau \\ rS(x_i)U(x_i)(T+\tau-x_i) & x_i \geq \tau \end{cases}$$

We don't need integrals since the survival functions support is $\mathbb{N}$.

To understand $U(x_i)$, we need to provide a clean description of the simulation stopping criteria. For i in $[n]$, let T_{iM} be the first time at which some number M of messages have gone extinct. Now, let

$T = \max\{T_{iM} : i \in [n]\}$

So, if you're a message, your odds of being censored increase as your message ID increases

---

12/22/2022

I don't want any methodological goof ups, so I'm gonna read a few chapters of this book "Survival Analysis: A Self-Learning Text" to make sure everything is solid.

Ch 1. Introduction to Survival Analysis

- Type of problems addressed
- The outcome variable
- "Censored data"
- What survival and hazard functions are.

I. What is survival analysis?

Interested in time until an event occurs.

We assume only one event is of designated interest. If more than one event is considered, this is a recurrent event or competing risk problem.

Time = survival time; event = failure

Survival analysis can be applied to many clinical and engineering applications and even to such issues as recidivism.

II. Censored Data

- We have some information about survival time, but we don't know the survival time exactly.

Causes of censoring are usually:

1) A person does not experience the event before the study ends.
2) A person is lost to follow-up during the study period
3) A person withdraws from the study.

From the book, we have the following table

Person	Survival time	Failed (1) / Censored (0)
A	5	1
B	12	0
C	3.5	0
D	8	0
E	6	0
F	3.5	1

Note that this data is all right-censored. Data can also be left-censored, but usually it's right censored.

Right-censored: True survival time $\geq$ observed survival time
Left-censored: True survival time $\leq$ observed survival time

Example of left-censored: If you enroll in a study and test positive for HIV, then the true infection time lies between the enrollment and test times.

There's interval censoring as well.

III. Terminology and Notation

T: Random variable for a person's survival time

t: Any specific value of T

d: a (0,1) random variable denoting failure (1) or censorship (0).

S(t): the survivor function

h(t): the hazard function

$S(t) = \mathbb{P}(T > t)$. Theoretically, S is a smooth function that:

- is nonincreasing
- satisfies $S(0) = 1$
- satisfies $\lim_{t \to \infty} S(t) = 0$.

Estimates are step functions (usually)

$$h(t) = \lim_{\Delta t \to 0} \frac{\mathbb{P}(t \le T < t + \Delta t \mid T \ge t)}{\Delta t}$$

- The hazard function h(t) gives the instantaneous potential per unit time for the event to occur, given that the individual has survived up to time t.
- Always nonnegative, with no upper bound.

- Constant hazard rate, $h(t) \equiv \lambda$, means the survival model is exponential.
- Increasing Weibull
- Decreasing Weibull
- Lognormal

Note that h:

- Is an instantaneous potential
- May be used to identify a specific model form that fits the data
- The survival model is usually written in terms of the hazard function

$$S(t) = \exp\left[-\int_0^t h(u)\,du\right]; \quad h(t) = -\frac{1}{S(t)} \frac{dS(t)}{dt}$$