ε is small enough, we'd never see an input from B-$\mathcal{B}_L$ in a ~~finite~~ feasible observation time, so we'll use $\varepsilon = 0$ in the implementation. But hey, this satisfies proportionality.

Finally, for independent censoring, that clause "conditional on exposure and covariates" hints that I need to make another appeal to the stationary distribution.

But now, I'm tired. It's time for bed.

12/5/22 I wasn't able to find my elaboration on what "conditional on exposure and covariates" means, but for the model at hand—let's call it the brancher-annihilator—I have an argument in mind:

If the Markov chain is ergodic, then the independent censoring condition is satisfied for the extinction time. Why? Well:

- The only reason a message would be censored is that the observation period ended. The extinction time is ultimately governed by the chain's stationary distribution, so the censored messages have the same survival prospects as any other message.
  - Admittedly, having an expression explicitly relating the extinction time of a message at node i to the stationary distribution of the brancher-annihilator (BrAn) would make this argument airtight.

With this in hand, we add a task to the queue:
4) Express survival time probabilities as a function of the stationary distribution

And now, we can move on to task 3

Biased? Negligibly ✓ Consistent? Sample size?

For the first two questions, the original paper

[2] Kaplan and Meier, "Nonparametric Estimation...", 1958

states in section 2 that the KM curve is <u>consistent</u> and of <u>negligible bias</u> (unless excessive averaging is done). They provide a correction for the bias, but note it is "neither feasible nor worthwhile."

They also talk about the variance, but I think I'd rather look at confidence intervals.

Recall that a $(1-\alpha)100\%$ confidence interval $[a, b]$ for an estimator $\hat{\theta}$ of a parameter θ is such that

$$P(\theta \in [\hat{\theta}-a, \hat{\theta}+b]) = 1-\alpha.$$

In basic stats, $\hat{\theta}$ was a point estimator, but in this case it's a function.

Aha, but something to keep in mind is that most of the literature is concerned with survival functions over $\mathbb{R}$, but I'm interested in a function over $\mathbb{N}$, so maybe I can resort to more basic methods.

Wikipedia points to three statistics that may serve my purpose.

- Pointwise confidence intervals
- The Hall-Wellner band
- The equal-precision band

Let's check out the pointwise confidence intervals first.

The paper is

[3] Fay, Brittain, Proschan, "Pointwise confidence intervals...", 2013

- They mention Greenwood's variance

Really quickly, I'd like to describe a better modification to the Markov chain.

Suppose the raw chain (that is, the one determined by $\mathrm{BrAn}(A, B)$) is not ergodic. We will construct a modified chain that is ergodic.

Let $C_1, \dots, C_m$ be the communicating classes of the raw chain. Let x_i^* be an arbitrary state in C_i, $i \in [m]$. The alphabet at each x_i^* will be $B \cup \{\bar{I}\}$, and the probability distribution will be $(1-\varepsilon)\mathrm{unif}(B)$ over B and $\varepsilon\,\mathrm{unif}(\bar{I})$ over $\{\bar{I}\}$.

Next, the alphabet at $\{\bar{0}\}$ will be $[1]_0^n = B \cup B^c$ and the distribution is $(1-\varepsilon)\mathrm{unif}(B)$ over B and $\varepsilon\,\mathrm{unif}(B)$ over B^c.

The modified chain is finite. It is aperiodic because $\bar{0}$ can map to itself. For irreducibility, fix starting state x_S and target state x_T. The trajectory from x_S will enter C_i (for some i) w/p 1, and then will hit x_i^* w/p 1. If the next 2 letters are $\bar{I}$ and x_T, then the resulting state is x_T. So every state communicates with every other state, ergo irreducibility, ergo ergodicity.

11/18/22 From [3] Cantor, "Projecting the standard error of the Kaplan-Meier estimator" (2001), take formula

$$\widehat{var}_2 = \hat{S}^2(t) \sum_{t_i \le t} \frac{d_i}{N(t_i)[N(t_i)-d_i]} = V_t$$

to project V_t, replace observed times t_i by a partition of the interval $[0, t]$ by $0 = x_0 < x_1 < \cdots < x_n = t$, with $x_i - x_{i-1} = \Delta x$. Now, d_i can be thought of as the number of deaths in the interval $[x_{i-1}, i)$, and $N(x_i)$ the number at risk at time x_i.

Replace each ~~approximation~~ value by expected value

$$d_i \to \lambda(x_i) N(x_i) \Delta x$$