Another thing to consider is what the ergodicity adjustment actually means. For an ε-BrAn(A, B) system, the smaller ε is, the slower the rate of convergence to the stationary distribution. So I may want to set ε to something non-negligible so that convergence is faster, maybe 10^{-3}?

Oh yes, also: tracking one message at a time means that the simulation runs longer, hence the initial state for a given message is more and more likely to be distributed according to the stationary distribution as the message number grows. This means that if someone manages to find an efficient way of calculating the stationary distribution, then these predictions can actually be tested.

1/15/2023

Try to understand independent censoring vs. random censoring.

Ex) Estimate the 3-year survival (from some disease) among those in group A. 100 individuals disease free for three years. Over the 3-year period, 20 contract disease. So

- The 3-year risk of disease for those in group A is estimated to be 0.20
- The estimated 3-year survival is 0.80.

Now, we continue the study for 2 years to estimate the 5-year survival for group A. Want to continue to follow the 80 individuals remaining, but 40 refuse (are censored). Of the 40 remaining, 5 contract the disease.

Time	# at risk	# events	# survived
0-3	100	20	80
3-5	40	5	35

Estimated survival? $\frac{80}{100} \times \frac{5}{40} = 1/40 = 0.025$? No

If we knew what happened to the 40 individuals who were censored, we could just sum the total number of events and the total number who survived.

Under an assumption of independent censoring or random censoring, we assume that the 40 censored individuals were similar the 40 who remained at risk in terms of their survival probabilities

5 of the 40 uncensored developed the disease, so we assume 5 of the 40 censored did too. So over 5 years. 20 developed the disease in the first 3 years. 5 in the 3-5 year period, and we estimate 5 in the censored group too. So $20+5+5=30$ are estimated to have contracted the disease. Thus, $100-30=70$ survived over the 5 year period. So the estimated survival is 0.70 under indep or random censoring.

These two assumptions both allow you to use the observed data to estimate the survival of the censored group.

Independent and random sampling are the same if there only are group (i.e., no predictor variables).

Previous example extended: Consider group B, 100 people. Disease free at start. Estimate their 5 year survival now

Time	# at risk		# events	# survived
0-3	100	10 censored	40	60
3-5	50		10	40

Under independent censoring. 20% of the censored had events, so 2 more had events. Hence the number of events is

$40+10\cdot 2=52$.

Hence survival is $48/100=0.48$

Combining both groups, 200 at risk, 60 had events (A 20, B 60) in 0-3 leaving 140 (A 80, B 80) surviving by year 3, and at year 3, 50 were censored (A 40, B 10).

A higher proportion of censoring occured in group A than group B: A: 40/80 = 0.50, B: 10/60 = 0.17, so censoring was not random.

However, conditional on each level of covariates, the censoring was random, so the censoring was independent b/c:

Independent censoring is random censoring conditional on each level of covariates.

If 40 of group A and 30 of group B were censored at the 3 year mark that would be random censoring.

In this example, we assumed independent censoring and then showed it doesn't imply random censoring. But how do you verify the assumption of independent censoring.

The third assumption is <u>non-informative censoring.</u>
Depends on the distributions of
(1) the time-to-event random variable, T, and
(2) the time-to-censorship random variable, C.

- Conceptualize the dsbn of T as dsbn of survival times if there is 0 censorship
- Conceptualize the dsbn of C by considering censorship times for all subjects who would <u>not</u> have had an event by the end of the study period.

Non-informative censoring occurs if the distribution of T provides no information about that of C, and v.v.

Ex: Indep and random but informative

Subject A gets event ⟹ subject B (randomly selected) gets event; e.g., family member of subject A leaves study.

Assume: censored subjects represent subjects at risk at any time
Then: indep & random, but informative.

Bias can occur if censoring is not independent; e.g., censoring people who experience side effects.

Independent censoring is the most relevant.

Ch. 2: Kaplan-Meier Survival Curves and the Log-Rank Test

- Will need to remember purpose of survival analysis, basic notation and terminology, and the basic data layout for the computer

Indiv	End Time t	Censored d	X_1	...	X_p

Kaplan-Meier curves.
Log-rank test

Recall the question you want to answer: What features of the network predict the survival time of a message? So I can regress on ~~network size~~ local metrics, global metrics, and also the activation at the time of injection.
I don't care about the load value, it just.

muddies the picture. Maybe I'll make $\beta = X_n$

I. Review

Kaplan-Meier analysis is based on the alternative data layout.