Title: Contingent probability methods for fitting survival models when covariates to survival vary throughout the survey period

Authors: Ana Miller-ter Kuile, Kiona Ogle, Jamie Sanderlin

Baby:

Survival analyses have broad application across x, y, z subject areas and are important for understanding many processes, from human health to forest management.

Zoomed in baby:

There are different analyses created for different types of survival analyses. Many traditional analyses assess survival across one survey interval, usually with one “treatment” applied to a group throughout that entire interval and either outcomes or time to outcomes are measured for different individuals/data points. Others, such as in wildlife biology, try to link survival to finer-scale covariate data that is likely to change throughout the survey period. To do this, most studies employ a logistic exposure model to survey data that takes place over multiple sub-intervals throughout the survey period for each individual being tracked in the population. Then, these models link daily survival values to survey period survival by exponentiating daily survival to the length of a survey interval. These methods are appealing for applications in ecological contexts because they account for uncertainty about when an individual entered the population and when within a later survey interval they die or change states (such as when a nest produces fledglings in a nest survival study). They also allow for covariates to be finer scale than the entire lifetime of the individual. So, for example, a nest being monitored for nest fate may experience lower temperatures early in the season and warmer temperatures later in the season. Modeling survival with an interval-level temperature variable rather than an average temperature for the overall nesting period will give a better understanding of how temperature shapes nesting ecology.

Werewolf:

While methods that allow covariates to change across survey intervals are broadly used

and account for some error in these interval survival processes, these models are often defined such that each interval for each individual is a data point in a Bernoulli process. As a result, these datasets can have an exaggerated number of survival (success) values (1) since individuals that die (fail) are not monitored after death (failure). While this may be less of an issue when the number of survey intervals is low or when there is a relatively equal number of surviving and dead individuals in the dataset, as survey interval number increases or the survival-failure numbers are skewed, these models may not properly predict the data distribution of the data generating the data. Thus, using these models as predictive tools may be limited and biased.

Silver Bullet:

In this study, we present survival analyses with three different approaches. The first approach does not allow covariate values to vary throughout the survey period. The second two are two different applications based on the principles of the logistic exposure model where covariates can vary throughout the survey period. First, we fit the traditional logistic exposure model, where the data come from the 1-0 values at each survey interval for each individual. Then, we fit a modified model in which data come from 1-0 values for the final fate of each individual in the dataset, but where the overall survival probability that describes the Bernoulli process is shaped by interval-specific survival probabilities and depends on the final fate of the individual in that period and how many total survey intervals for each individual. We first fit these three models to simulated datasets with varying skewness in 1-0 values due to a combination of increased number of survey intervals per individual and skewness in final 1-0 outcomes. We then compare the models for an empirical dataset where the final outcome for individuals is skewed (~75% success) and each individual was surveyed for an average of 5 survey intervals (min 1, max 15). Thus, 1-0 distribution is skewed both throughout the survey period (~2:1 successes:failures) and even more skewed for interval-specific fates (~16:1 successes:failures). We present our examples in a Bayesian framework to account for parameter uncertainty and to aid in fitting custom probabilities. We expect that when there are not covariates that vary throughout the survey period, survey-long survival is a sufficient approach. However, when conditions vary, we expect that survey-long survival studies will not detect important variables that may change throughout the survey period. When covariates change throughout the survey period, we expect that studies where there are relatively few survey intervals per individual, regardless of overall distribution of final fates, traditional logistic exposure approach should generate models that are predictive of input data distributions. However, we expect that as the number of survey intervals per individual increases, especially when the distribution of 1-0 outcomes is skewed, that custom probabilities will generate a more predictive model of overall data distributions. Regardless, we expect that parameter estimates will be similar between these two interval-specific approaches, so the choice of which method to use comes when models are intended to be predictive as opposed to exploratory.

Simulations:

Combinations of survey interval number per individual and the ratio of final 1-0 values

1. Many surveys, unequal 1-0 (very poor fit) – this is our dataset
2. Many surveys, equal 1-0 (some poor fit)
3. Few surveys, unequal 1-0 (should fit okay)
4. Few surveys, equal 1-0 (should fit well)