**Grand Framework**

At its core, GenEst is an elaboration of a binomial probability model , where is the observed number of carcasses and is the detection probability. If is known, then is an unbiased estimator for . In a slightly more complicated scenario in which total mortality is split into two groups, and , then is unbiased for . GenEst makes extensive use of this simple idea of splitting the carcass observation data into distinct subgroups, estimating mortality in each subgroup, and combining subgroups into larger groups by summing. However, a number of technical difficulties must be overcome to make this simple idea work as a complete estimator that produces accurate confidence intervals, including accounting for the dependence of on the time of carcass arrival, estimating as a function of covariates, characterizing the uncertainty in , accounting for correlation structure of the 's among various subgroups, accounting for binomial uncertainty about , and accounting for unsearched area. GenEst provides solutions for each of these difficulties.

The technical details are lengthy but are explained in full in sections xxx-xxx.

**Splitting Mortality Estimates by Carcass and Recombining into Subgroups**

For each carcass, , that is discovered in carcass surveys, roughly speaking its contribution to the estimated total mortality is the reciprocal of it inclusion probability, , but with two types of uncertainty that we account for via parametric bootstrapping: (1) uncertainty associated with estimating (sections xxx-xxx), and (2) the binomial uncertainty associated with estimating (section xxx). The result is stored in an matrix of carcass contributions to the estimated total mortality

Each column of represents a single realization of the simulated carcass contributions to estimated mortality. Uncertainty is captured in the distinctions among columns.

Mortality estimation by carcass group is accomplished by subsetting the rows and then taking column sums. For example, total mortality (no subsetting) is estimated by taking column sums over all the carcasses to create a vector of estimates, , from a parametric bootstrap from the sampling distribution of . The matrix can be subsetted to estimate mortality among desired subgroups in a similar way. For example, to estimate mortality for species B, calculate , where indicates that carcass belongs to species B. GenEst extracts sample statistics (like median and confidence intervals) from these empirical vectors to summarize mortality estimates for subgroups or "splits" as defined by the user.

**Temporal Splits**

GenEst can also split mortality estimates by user-specified time intervals or by temporal covariates like season. To split the matrix by carcass arrival times, GenEst relies on the matrix of simulated arrival intervals, , to construct an array of mortality estimates as:

,

where indicates that the simulated arrival time of carcass is within the specified interval .

To understand the structure of the matrix, first note that detection probability for a carcass depends in part on its arrival time. For example, if monitoring runs from April 1 through October 31, carcasses arriving near the end of October may be available for discovery for only one or two searches, while carcasses arriving in April may be available for many searches. Also, search conditions may vary with season. However, although carcass discovery times are recorded, arrival times are not known. Discovered carcasses may have arrived some time after the previous search or in any search interval prior to that. Even though arrival times cannot be known with certainty, GenEst analyzes the search data to derive probability distributions of arrival times for each carcass and, from these, creates an matrix of simulates arrival intervals, . For example, suppose and the carcasses were discovered on searches and of a monitoring period with . The arrival matrix, , might look like the following:

The simulated arrival intervals for the first carcass are identically 1 because the carcass was discovered on the first seach (at ) and is assumed to have arrived in the first search interval, . In practice, the assumption that all carcasses discovered in carcass surveys arrived after is often enforced by disregarding carcasses found in a careful clean-out search at . The second carcass was discovered in search 3. According to the arrival probabilities, the carcass was more likely to have arrived in interval 3 than any other interval, but there is a chance it arrived as early as the first interval, as reflected in the row of simulated arrivals. The third carcass was found on the 12th search and, in theory, could have arrived in any interval prior to its discovery. However, it is highly unlikely that it arrived more than a few intervals prior to the 12th because, if it had, chances are that it would have been removed by scavengers or previously found by searchers.

**Estimation of Arrival Probabilities**

Suppose searches are conducted at times, . For each carcass discovered in the carcass surveys , define as the event that a the carcass was observed during search (at time ) and as the event that the carcass arrived in interval . Then, the probability that the carcass arrived in interval is:

In theory, is positive for , but, in practice, decreases rapidly with , so in most cases only the first few terms need to be calculated. The default for the GUI is to calculate arrival probabilities for up to 8 intervals prior to discovery; the command-line user may override the default in functions estM and estg. GenEst makes a further assumption that the arrival rate is constant in the search intervals preceding carcass arrival, so

where length of interval .

The estimator is robust to variation in arrival rates among different search intervals provided that there is not an abrupt change in arrival rate from one interval to the next. However, even in these unusual conditions, the potential for bias in estimating total would be small because the same pattern would need to recur with a large fraction of the carcasses. The potential for bias would be further attenuated by high searcher efficiency or low persistence times.

Calculation of is based on calculation of :

where is the probability that a carcass persists days after arrival (section xxx), and are searcher efficiency parameters (section xxx), (the index on the product) runs from the search interval immediately before carcass discovery to the first search interval of the monitored period. A carcass may arrive in one season and be discovered in a later season after search conditions have changed. To account for this possibility, in the calculation of a carcass's for a given , the carcass is assigned the search characteristics appropropriate to the assumed arrival interval, . For example, if a carcass is discovered on the first search in the fall, and search conditions (that is, searcher efficiency and carcass persistence parameters) change from summer to fall, then is calculated with the fall SE and CP parameters and is calculated with the summer parameters.

Arrival-interval probabilities, , are calculated for each carcass and search interval, and an integer-valued matrix of simulated arrival intervals, , is constructed. Each column represents the arrival interval of each carcass in a simulated "year" or set of carcass arrivals. As mentioned previously, carcass detection probabilities depend on arrival times, which are given in the arrival matrix *.* For each simulated year (column in ), the detection probability for each carcass is estimated as

to construct an matrix, , of carcass detection probabilities.

**Uncertainty in Estimating**

In the unlikely case that is identical for all carcasses and is known exactly, there may still be a great deal of uncertainty about the true mortality due to binomial variation. GenEst accounts for that uncertainty in a novel way that lends itself well to the splits framework.

We can assume , but, unfortunately, we don't know . We can, however, estimate by and (almost) define a new random variable to account for the variation in . The uncertainty in would then be accounted for in . The problem is that need not be an integer and cannot serve as the index for a binomial random variable. Instead, we define , where is a continuous generalization of the standard binomial distribution (Ilienko 2013), which is implemented in the R package cbinom (Dalthorp 2018). The continuous binomial spreads the binomial distribution's probability mass on each integer to the interval , so the distribution is a smoothed version of the binomial but shifted slightly to the right. The mean of the continuous binomial distribution with parameters and is , where is the CDF of the continuous binomial:

and

is the incomplete beta function. The expectation of a random variable is

which is approximately and can be calculated numerically. Since the expected value of a binomial random variable is , exceeds by about 0.5 on average. Thus, and would be biased for estimating , so we subtract off the bias before dividing by :

Because this estimator is unbiased, mortality estimates can be summed. For example, if is the estimated mortality in area A and is the estimated mortality in area B, then is an unbiased estimator for the two areas combined. There is nothing special about "area" here; A and B could represent times, species, carcass sizes, search teams, turbine types, or other variables. GenEst takes this idea to the limit and builds the matrix, , in which each carcass represents its own "area". Users then define how they wish to split the overall mortality into summary groups. In the simplest case, total mortality is estimated as the sum of each carcass's contribution to the total, that is, the column sums of the matrix.

**Accounting for Unsearched Area**

Unless the user is interested in estimating mortality strictly in the searched area, values must be adjusted to account for unsearched area. In practice, some carcasses are likely to fall outside the searched area in a given unit (e.g., in an unsearched part of a search ring at a solar power tower facility or beyond the search radius at a wind turbine). In addition, there may be units at a site that are not searched at all. For each search unit, , the expected fraction of carcasses arriving outside the searched area is the *density-weighted proportion* or (Huso and Dalthorp, 2014). The expected fraction of carcasses arriving outside the units searched at a site is the *sampling fraction* or *f*. Thus, the contributions of the respective carcasses to the estimate of total mortality are then summarized in a array as , where is the diagonal matrix of values associated with the carcasses. For example, if carcass 1 was found at unit 17 and was a medium-sized bird, then would be the for medium-sized birds at unit 17. would then be calculated based on this adjusted .

## Searcher Efficiency

Let be the initial searcher efficiency for fresh carcasses, or, more precisely, the conditional probability of detecting a carcass on the first search after carcass arrived, given that the carcass is present at the time of the search. Let be the fractional change in searcher efficiency with each successive search. Then, if searcher efficiency trial carcasses that are missed in one search are left in the field for possible discovery on later searches (Appendix A1), and can be estimated simultaneously as functions of categorical covariates.

Specifically, let be the probability that carcass is found during the first search after carcass arrival given that it is present at the time of the search. The model allows to depend

on a vector of covariates as , where is a vector of coefficients associated with combinations of covariate levels. The model allows a constant multiplicative reduction in detection probability in subsequent searches. The probability of finding carcass *i* during search *j* is

|  |  |  |
| --- | --- | --- |
|  |  | (A1) |

where may also be a logit-linear function of covariates and coefficients. These covariates need not be the same as those used to model

Consider the search outcome history for a given carcass. If the carcass was scavenged before the first search occasion, then the carcass provides no information to estimate searcher efficiency, and the model ignores these carcasses. If the carcass is present during one or more search occasions, then the probability of its search history is determined by that history and (A1). It is easiest to see with a couple of examples. Suppose carcass (from the searcher efficiency field trials) was missed on three searches and detected on the fourth search. Its search history is then (0, 0, 0, 1) (in the notation of the data), and the probability of this search history is

where denotes the contribution of carcass to the joint likelihood.

A carcass may never be found. For example, suppose carcass was missed on the first two search occasions and was then scavenged before the third search. This carcass has search history (0, 0), and

The log-likelihood of the data is then expressed as

where denotes the total number of carcasses. Let denote the number of zeros in the search history for carcass *i* (that is, is the number of times carcass was missed in searches), and let denote the search occasion where the carcass was found, with if the carcass was never found. The, will then have the form

where if and zero otherwise. Thus the log-likelihood depends on the data only through the number of missed searches and whether or not the carcass was eventually found.

The log-likelihood, , of the data is numerically maximized in using R's optim function. From the theory of maximum likelihood, the vector of maximum likelihood estimators (MLEs) for the parameters (is asymptotically multivariate normal with mean equal to the true parameter values and variance-covariance matrix given by the inverse information matrix, which is returned by optim as the hessian.

To account for the uncerainty in estimating and parameters, GenEst first approximates the sampling distribution of by simulating from

where is the Hessian matrix returned by optim(). Simulated sampling distributions of and are obtained by back transformation of simulated . Specifically, let represent the covariate vector for the th carcass, then

where and are the components of associated with and , respectively.

Carcass persistence times are modeled using censored exponential, Weibull, lognormal, and loglogistic survival models, which are fit by maximum likelihood estimation using the R functions survival::survreg (Therneau 2015) and optim (R Core Team, 2018). Both the location and scale parameters (Kalbfleisch and Prentice 2002, chapter 2) may depend on categorical covariates, such as visibility class, season, or other factor. As with the searcher efficiency estimates, vectors of simulated persistence parameters are generated from the fitted models.

Dalthorp, D. H., J. Simonis, L. Madsen, P. Rabie, J. Stuydevant, F. Korner-Nievergelt, R. Bispo, R. Wolpert, and M. M. P. Huso. 2018. Generalized Estimator of Mortality (v1.0) software user

guide. USGS Data Series XXXX. USGS, Corvallis, OR.

Ilienko, Andreii (2013). Continuous counterparts of Poisson and binomial distributions and their properties. Annales Univ. Sci. Budapest., Sect. Comp. 39: 137-147.

Kalbfleisch, J. D. and Prentice, R. L., The statistical analysis of failure time data, Wiley, 2002.

R Core Team (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Therneau, T., 2015, A Package for Survival Analysis in S: The Comprehensive R Archive Network Web page, version 2.38, http://CRAN.R-project.org/package=survival.