Assignment 1

Sam Olson

Preface

For each of the following situations, define random variables that might be appropriate for conducting an analysis. If distributions are to be assigned to these variables, what should be the support of those distributions? Would it be reasonable to consider the variables to be independent? In some cases there may not be a single correct answer. While it is certainly reasonable to be thinking about some type of model you are not being asked to identify a model structure that you would use to address the objectives identified. You are only being asked to define appropriate random variables (and perhaps covariates), identify what support a distribution assigned to them should have, and make a preliminary assessment about an assumption of independence. Do not suggest a model, you will loose points if you do.

Q1

1. (20 pt.) As part of its Long Term Resource Monitoring Program, the USGS Upper Midwest Environmental Sciences Center located in LaCrosse, Wisconsin has sampled sediment from the Upper Mississippi River from about 120 sites at each of six reaches (or stretches) of the river, from 1992 to 2004. In each stretch of the river, samples were taken from a number of primary habitat categories used as strata in a sampling design. Those categories were Backwater, Impounded Water, Side Channel, and Main Channel Border. Sampling locations were selected separately each year so that repeated sampling of the same location over time did not occur. The sediment samples are brought back to the laboratory, run through a sieve, and the types and numbers of invertebrates are recorded, as is the specific predominant sediment type of sand, silt,or clay (there are actually 6 categories, but 3 is enough to get the idea). Water depth is also measured at the time each sample is collected.

The USGS would like to use these data to address a number of questions related to the status of mayflies (Ephemeroptera) in the river. Mayflies form the basis for a number of aquatic food chains and are also generally considered an indicator of water quality (rather, their absence is considered an indicator of a lack of water quality). There are any number of objectives that might be identified for this study. Here, we will be concerned with only two. Restrict your answer to issues that are relevant for these two objectives.

(a)

Is the presence/absence of mayflies at sampling locations related to the primary habitat category and/or the specific sediment type?

(b)

Has the abundance of mayflies exhibited a systematic change over the period 1992 to 2004?

Answer

Define appropriate random variables (and perhaps covariates), and what support a distribution assigned to them should have:

A R.V. for the presence of mayflies (in a sample), $Y_1 : \Omega \to \{0, 1\}, \quad Y_1(\omega) \in \{0, 1\}$

A R.V. for the number of mayflies (in a sample), $Y_2 : \Omega \to \mathbb{N}_{\not\sim}, \quad Y_2(\omega) \in \mathbb{N}_{\not\sim}$

(Covariate) A R.V. for the primary habitat category, $X_1 : \Omega \to \{\text{Backwater}, \text{Impounded Water}, \text{Side Channel}, \text{Main Channel B}$
$\{\text{Backwater}, \text{Impounded Water}, \text{Side Channel}, \text{Main Channel Border}\}$

(Covariate) A R.V. for the sediment type, $X_2 : \Omega \to \{\text{Sand, Silt, Clay}\}, \quad X_2(\omega) \in \{\text{Sand, Silt, Clay, ...}\}$ ("3 is enough to get the idea")

(Covariate) A R.V. for year, $X_3 : \Omega \to \{1992, \ldots, 2004\}, \quad X_3(\omega) \in \{1992, \ldots, 2004\}$

Preliminary assessment about an assumption of independence:

Independence of the presence of mayflies across samples *within a year* I would say seems unlikely, since nearby sampling locations (though separate) may share similar morphological characteristics, especially since (as I believe) mayflies are born in clusters of eggs (a clutch?, seems to depend on the specie of mayfly).

Also, and w/r/t *between years*, I believe independence is reasonable given the "repeated sampling" method of not repeating the exact same location, though again, independence *within a year* I'd argue is likely not being met.

Q2

2. (10 pt.) A study was conducted at the College of Veterinary Medicine at ISU to examine the efficacy of a vaccine for Porcine Reproductive and Respiratory Syndrome (PRRS) virus. This virus infects sows but affects piglets. In the study, 12 pregnant adult sows were given the vaccine and another 12 were not. Both groups were "challenged" (i.e., exposed to the virus). Sows were housed separately. The number of piglets born normal, born weak, and born still born were recorded for each sow. The objective of analysis was to determine whether the vaccine was effective in reducing the effects of the PRSS virus.

Answer

Define appropriate random variables (and perhaps covariates), and what support a distribution assigned to them should have:

A R.V. for the number of normal piglets born to sow i, $Y_1 : \Omega \to \mathbb{N}_0, \quad Y_1(\omega) \in \mathbb{N}_0$

A R.V. for the number of weak piglets born to sow i, $Y_2 : \Omega \to \mathbb{N}_0, \quad Y_2(\omega) \in \mathbb{N}_0$

A R.V. for the number of stillborn piglets born to sow i, $Y_3 : \Omega \to \mathbb{N}_0, \quad Y_3(\omega) \in \mathbb{N}_0$

(Covariate) A R.V. for vaccination status of sow i, $X_1 : \Omega \to \{0, 1\}, \quad X_1(\omega) \in \{0, 1\}$, where "1" = vaccinated, "0" = not vaccinated.

Preliminary assessment about an assumption of independence:

Independence *between sows* seems reasonable, since they are housed separately and treated as experimental units, i.e., we have separation of units at level at which the "treatment" of vaccination is applied. Independence *within a litter* (across piglets) is likely not a reasonable assumption though, since piglets share the same vaccination status of their sow (and possibly other biological factors, genetics, etc.).

Q3

3. (10 pt.) A problem considered by Dr. Dixon about 15 years ago was the topic of a seminar he presented to the department. This problem involved the capture of insects by a predatory plant species, a member of the family of pitcher plants Sarraceniaceae. These plants have a long central tube with a hood-shaped part at the upper end. The tube has hair-like structures that point downward. Insects that enter the tube are not able to move back up because of these hairs, and eventually are digested by enzymes in the plant. Insects that do not enter the tube may obtain nectar from the plant without becoming plant food. A primary prey species of the pitcher plant species involved in this study are a certain type of small wasp. The study was designed to determine how effective these plants are at capturing wasps.

The study consisted of two parts. The first part involved direct observation of the plants for several hundred hours. The data recorded were the number of wasps visiting the plants and the number of these that were captured by the plants. There were a total of 376 "plant-hours" of observation, 157 visits, and 2 captures. The second part of the study involved cleaning out a number of plants, leaving the study site, and returning about 2 weeks later (it takes the plants longer than 2 weeks to totally digest a wasp that is captured). The data recorded in this part of the study were the number of wasps captured by the plants over a period of 2 weeks, which was equivalent to 1416 "plant-hours". There were a total of 6 wasps captured in this indirect observation portion of the study.

Our concern here is not with the actual numbers that resulted from this study, but rather with defining random variables that might be used in a statistical analysis. The focus of the seminar by Dr. Dixon was how information from the indirect observation part of the study could be combined with information from the direct observation part of the study to improve estimation of the rate of visits by wasps and the probability of capture given a visit, which were the objectives of the study.

Answer

Define appropriate random variables (and perhaps covariates), and what support a distribution assigned to them should have:

A R.V. for the number of wasp visits to plant p during an observed plant–hour, $V : \Omega_{\text{PH}} \to \mathbb{N}_0, \quad V(\omega) \in \mathbb{N}_0$

A R.V. for the number of captures during an observed plant–hour, $C : \Omega_{\text{PH}} \to \mathbb{N}_0, \quad C(\omega) \in \mathbb{N}_0$

(Alternatively) A R.V. for capture during an observed plant-hour (during a visit), $Z : \Omega_{\text{visit}} \to \{0, 1\}, \quad Z(\omega) \in \{0, 1\}$

A R.V. for the number of captures accumulated over an indirectly observed plant–window, $K : \Omega_{\text{PW}} \to \mathbb{N}_0, \quad K(\omega) \in \mathbb{N}_0$

(Covariate) A R.V. for observation mode, $X_1 : \{\Omega_{\text{PH}}, \Omega_{\text{PW}}\} \to \{\text{direct}, \text{indirect}\}, \quad X_1(\omega) \in \{\text{direct}, \text{indirect}\}$

Unsure if these should be added (Covariate) A R.V. for plant identifier, $X_2 : \{\Omega_{\text{PH}}, \Omega_{\text{PW}}\} \to \{\text{plant IDs}\}, \quad X_2(\omega) \in \{\text{plant ID 1, plant ID 2, ...}\}$ (number of plants not specified)

(Covariate) A R.V. for exposure time (hours observed for that unit), $X_3 : \{\Omega_{\text{PH}}, \Omega_{\text{PW}}\} \to [0, \infty), \quad X_3(\omega) \in [0, \infty)$

Preliminary assessment about an assumption of independence:

Independence *between different plants* seems reasonable, though independence *within the same plant* (across plant-hours) may be unreasonable. This assumes that plants are spatially separated, e.g., plants don't share a pot/plot. My reasoning for a lack of independence *across plant–hours for the same plant* is due to "expected temporal clustering of wasp activity". At the visit level, capture(s) within the same plant (and closer in time) are unlikely to be independent because visits share the same plant (and other potential characteristics of interest). Also, in Phase 2, the indirect-window counts K may be dependent on direct plant–hour quantities

(number of wasp visits, number of captures) if they correspond to overlapping plants/times; for disjoint plants/times (different plants at different time periods), independence seems plausible.