Predator Preferences Model

S. Bonner and E. A. Roualdes

May 27, 2014

1 Generic Model

Strauss' statistic L_{st} [Str79] is the difference between r_{st} the proportion of prey species s found in the gut of a predator during occurrence t and p_{st} the proportion of prey species s found in the habitat of said predator during occurrence t. When the statistic $L_{st} := r_{st} - p_{st}$ is equal to zero we say the predator ate prey species s randomly during time t. A positive difference, $L_{st} > 0$, indicates preferential eating of prey species s, and negative values indicate aversion to prey species s at time t. We can replicate and add to the modeling considered by Strauss with the following set up.

Let $X_{jst} \sim \mathcal{P}(\lambda_{st})$ denote the number of prey species s found in the gut of predator j during occurrence t; $j \in \{1, \ldots, J\}, s \in \{1, \ldots, S\}, t \in \{1, \ldots, T\}$. For now we ignore the fact that we only observe a 1 if predator j ate prey species s during occurrence t or a zero if the predator did not. Further, $Y_{ist} \sim \mathcal{P}(\gamma_{st})$ will denote the number of prey species s found in trap i, $i \in \{1, \ldots, I\}$, hypothesized to represent the habitat of predator j, during time occurrence t. Interest lies in the null hypothesis

$$H_0 : \boldsymbol{\lambda}_t = c\boldsymbol{\gamma}_t, \forall t \tag{1}$$

where $c \in \mathbb{R}$. This hypothesis suggests that the predator's eating preferences are independent of time, i.e. the ratio of the rate at which the predator eats prey species s to the rate at which prey species s is found in the habitat is constant across time. Strauss' L is a special case of this framework where only a particular combination of prey species s and time t is considered, and c is equal to the ratio of total number of prey found in the gut of the predator to the total number of prey found in the habitat.

We first consider the parameters of this model under the null hypothesis. Here, $S \cdot T + 1$ parameters will be estimated, γ_{st} for $(s,t) \in \{S \times T\}$ and c.

$$\hat{\gamma}_{st} = \frac{X_{\cdot st} + Y_{\cdot st}}{I \left(\frac{\sum_{st} X_{\cdot st}}{\sum_{st} Y_{\cdot st}} + 1 \right)} \quad \text{and} \quad \hat{c} = \frac{I \sum_{s,t} X_{\cdot st}}{J \sum_{s,t} Y_{\cdot st}}.$$

The parameters under the general alternative hypothesis, $H_1 : \boldsymbol{\lambda}_t \neq c\boldsymbol{\gamma}_t, \forall t$, are simply the maximum likelihood estimates.

$$\hat{\lambda}_{st} = \frac{X_{\cdot st}}{J} \quad \text{and} \quad \hat{\gamma}_{st} = \frac{Y_{\cdot st}}{I}$$

Null hypothesis (1) above is tested with the standard likelihood ratio statistic

$$\Lambda(X,Y) := -2\log \frac{\sup L(\theta_0|X,Y)}{\sup L(\theta_1|X,Y)} \tag{2}$$

where $\theta_0 = \{\boldsymbol{\lambda}_t = c\boldsymbol{\gamma}_t, \forall t : (\gamma_{st}, c) \in \mathbb{R}^{ST+1}\}$ represents the likelihood under the null hypothesis, and $\theta_1 = \{\boldsymbol{\lambda}_{st} \neq c\boldsymbol{\gamma}_{st} : (\lambda_{st}, \gamma_{st}) \in \mathbb{R}^{2ST}\}$ represents the likelihood under the alternative hypothesis. Then $\Lambda \dot{\sim} \chi^2_{ST-1}$ and null hypothesis (1) is rejected when $P(\chi^2_{ST-1} > \Lambda) < \alpha$.

1.1   Non-Count Gut Data

With smaller animals it is not always possible to determine a count of prey species in the gut of the predator species. In this case, we can treat the observations as missing and maximize the likelihood via the EM algorithm. Here however, the alternative hypothesis mentioned above, $H_1 : \boldsymbol{\lambda}_t \neq c\boldsymbol{\gamma}_t, \forall t$ is too general. We simplify it to

$$H_1 : \boldsymbol{\lambda}_t = c_t\boldsymbol{\gamma}_t, \forall t \tag{3}$$

so that $\boldsymbol{\gamma}_t$ can then provide some information about the unobserved counts produced by $\boldsymbol{\lambda}_t$.

Hence, we can still calculate $\Lambda(X,Y)$, but we do so using the EM algorithm as follows. Write out the complete log-likelihood.

$$l_{comp} = -J\sum_{st}\lambda_{st} + \sum_{st}X_{\cdot st}\log\lambda_{st} - I\sum_{st}\gamma_{st} + \sum_{st}Y_{\cdot st}\log\gamma_{st} + constant \tag{4}$$

Take the expectation with respect to the distribution of the missing data given the observed data and parameters, $f_{X|Y,\boldsymbol{\lambda},\boldsymbol{\gamma}}(x)$,

$$\mathbb{E}_{X|Y,\boldsymbol{\lambda},\boldsymbol{\gamma}}l_{comp} = -J\sum_{st}\lambda_{st} + \sum_{st}(\mathbb{E}X_{\cdot st})\log\lambda_{st} - I\sum_{st}\gamma_{st} + \sum_{st}Y_{\cdot st}\log\gamma_{st} + const$$

which is then maximized with respect to the parameters $\gamma_{st}, \lambda_{st}, c$, just as above since here $X_{\cdot st}$ is simply replaced by its expectation.

1.2   Unequal Trap Schedules

Consider a situtation where the traps (catching prey species) were left out for a differing length of time within each occurrence, e.g. trap i in month t was left out for 6 days, but trap i' in month t was left out for 3 days. We expect to catch an unequal number of prey in each trap simply because one trap was left out for a longer time within month t.

Put $ND_t := \sum_i numDays_{it}$ to be the total number of days all traps were left out during month t. Then the complete data log-likelihood under this scenario is as follows.

$$l = -J\sum_{st}\lambda_{st} + \sum_{st}X_{\cdot st}\log\lambda_{st} - \sum_t ND_t\sum_s\gamma_{st} + \sum_{st}Y_{\cdot st}\log\gamma_{st} + constant$$

2

Under null hypothesis (1) we maximize the following function over its agruments $(c, \gamma_{st}), \forall s, t$

$$l_0(c, \gamma_{st}) = -Jc \sum_{st} \gamma_{st} + \log c \sum_{st} X_{\cdot st} + \sum_{st} X_{\cdot st} \log \gamma_{st} - \sum_{st} ND_t \gamma_{st} + \sum_{st} Y_{\cdot st} \log \gamma_{st} + const.$$

And under alternative hypothesis (3) where c now depends on time, maximize

$$l_1(c_t, \gamma_{st}) = -J \sum_{st} c_t \gamma_{st} + \sum_{st} X_{\cdot st} \log c_t + \sum_{st} X_{\cdot st} \log \gamma_{st} - \sum_{st} ND_t \gamma_{st} + \sum_{st} Y_{\cdot st} \log \gamma_{st} + const,$$

for $(c_t, \gamma_{st}), \forall s, t$. The EM algorithm from above can be incorporated into this scenario exactly as before, keeping in mind the M-step involves maximizing these functions, which I can't seem to do analytically.

2   Simulations

In order to simulate and test data, we need to install an `R` package named `spiders`. To install the package `spiders`, you can use `devtools`, or download the files from the `GitHub` webpage.

```
## library(devtools) install_github('spiders',
## 'roualdes')
library(spiders)
```

We can then simulate some data by using the function `spiders::simData`. Input to this function includes the number of predator species observed at each time point, the number of traps used at each time point, the number of prey species of interest, the number of time points for which measurements were taken, and rate parameters $\lambda_{st}, \gamma_{st}$ for which the predator ate and observed, respectively, prey species s at time t. For now, we assume that the number of predators and traps is fixed across time points. Below we specify rate parameters that will be proportional across time; we expect to fail to reject null hypothesis (1) here.

```
Predators <- 20
Traps <- 20
PreySpecies <- 3
Times <- 12
ST <- Times * PreySpecies
l <- matrix(1:ST, nrow = Times, ncol = PreySpecies)
g <- matrix(2 * (1:ST), nrow = Times, ncol = PreySpecies)
```

Simulate some data as follows.

```
fdata <- simPref(PreySpecies, Times, Predators, Traps, l,
    g)
```

Then we can fit the model and calculate a p-value from these data. The output of the model fit is below.

```
(prefs <- predPref(fdata$eaten, fdata$caught))

$loglikH1
[1] 218220

$loglikH0
[1] 218202

$`-2log(LRT)`
[1] 36.21

$df
[1] 35

$p.value
[1] 0.4121
```

An example where we reject the null hypothesis would have rate parameters that are differently proportional across time. Consider the following.

```
l <- matrix(1:ST, nrow = Times, ncol = PreySpecies)
g <- matrix(2 * (ST:1), nrow = Times, ncol = PreySpecies)
fdata <- simPref(PreySpecies, Times, Predators, Traps, l,
    g)
predPref(fdata$eaten, fdata$caught)$p.value

[1] 0
```

References

[Str79] Richard E Strauss, *Reliability estimates for ivlev's electivity index, the forage ratio, and a proposed linear index of food selection*, Transactions of the American Fisheries Society **108** (1979), no. 4, 344–352.