

# Modeling Predator Preferences

Edward A. Roualdes and Simon Bonner

University of Kentucky

2014-09-10

## Abstract

Words

## Introduction

## Methods

## Data

We assume data are collected in the following manner. Traps are dispersed, for  $T$  time periods, throughout the habitat of the predator and prey of interest. Prey species, indexed by  $s \in \{1, \dots, S\}$ , are collected in the traps and counted at each time period. We assume these counts represent the number of prey species  $s$  the predator will encounter on average during time period  $t \in \{1, \dots, T\}$ . Hence, the counts of prey species  $s$  caught during time period  $t$  are hypothesized to be independent draws from a Poisson distribution with rate parameter  $\gamma_{st}$ . We further assume that the number of prey species found in the gut of the trapped predators, also follows a Poisson distribution with rate  $\lambda_{st}$ . Here, the parameter  $\lambda_{st}$  represents the rate at which the predator ate the encountered prey species  $s$  during time period  $t$ . We then make statistical claims about the relative rates that the  $S$  prey species are eaten to the rate at which they are encountered over  $T$  time periods.

Let  $X_{jst} \stackrel{\text{iid}}{\sim} \mathcal{P}(\lambda_{st})$  denote the number of prey species  $s$  that predator  $j$  ate during occurrence  $t$  where  $j \in \{1, \dots, J\}$ . Let  $Y_{ist} \stackrel{\text{iid}}{\sim} \mathcal{P}(\gamma_{st})$  denote the number of prey species  $s$  found in trap  $i$  during occurrence  $t$ ,  $i \in \{1, \dots, I\}$ . We make formal statistical statements about the relative magnitudes of the parameters  $\lambda$  and  $\gamma$ .

## Assumptions

Our data collection scheme assumes the following: 1) traps independently catch prey that are encountered at a constant rate, 2) all traps are equally likely to catch the prey species of interest, 3) predators eat, at a constant rate, prey species independently. Though there

Species		Time	
		constant	varies
	constant	$c$	$c_t$
	varies	$c_s$	$c_{st}$

Figure 1: The four hypotheses considered are shown by their symbolic representations, highlighting which indices are allowed to vary. This is essentially the range of the mapping  $\xi$ .

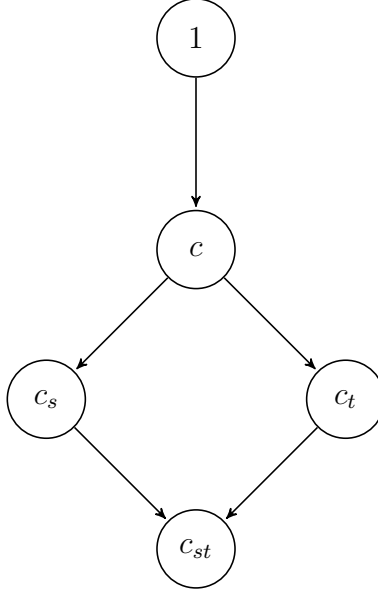


Figure 2: Hierarchy of hypotheses.

exists a non-zero chance that these assumptions are broken, we feel that slight deviations from them will not overtly effect the estimates.

We consider five hypotheses to provide variations on the relative magnitude of  $\lambda_{st}/\gamma_{st} = \xi(c_{st})$ , for a mapping  $\xi : c_{st} \mapsto \{1, c, c_s, c_t, c_{st}\}$ . When  $\xi$  maps to a constant with respect to one of the indexing variables,  $s$  or  $t$ , intuitive statements are realized about the predator's preferences. Table 1 depicts the ways in which the proportionality constants can vary, or not, for each of the four hypotheses. Figure 2 represents the order in which hypotheses can be tested against each other; how best phrase this, does the young Skywalker?

## Fully Observed Data

The likelihood function that allows for estimation of these parameters is as follows. Since we assume  $X_{jst} \perp Y_{ist}$  we can simply multiply the respective Poisson probability density functions together, and then form products over all  $s, t$  to get the likelihood

$$L(x_{jst}, y_{ist} | \lambda, \gamma) = \prod_{t=1}^T \prod_{s=1}^S \left\{ \prod_{j=1}^{J_t} f_X(x_{jst} | \lambda) \prod_{i=1}^{I_t} f_Y(y_{ist} | \gamma) \right\}. \quad (1)$$

In some cases analytic solutions for the maximum likelihood estimates are available, namely when the data are balanced  $J_t = J$ ,  $I_t = I$ , and  $\xi(c_{st}) = c$  or when estimating  $\lambda_{st}, \gamma_{st}$  under the hypothesis represented by  $c_{st}$ . In all other cases, analytic solutions are not readily available. In such cases, we rely on the fact that the log-likelihood  $l(\boldsymbol{\lambda}, \boldsymbol{\gamma}) = \log L$  is concave and iteratively solve

$$\hat{c} = \frac{\sum_{s,t} X_{.st}}{\sum_t J_t \sum_s \gamma_{st}}, \quad \hat{c}_t = \frac{\sum_s X_{.st}}{J_t \sum_s \gamma_{st}}, \text{ or } \hat{c}_s = \frac{\sum_t X_{.st}}{\sum_t J_t \gamma_{st}}, \text{ and } \hat{\gamma}_{st} = \frac{X_{.st} + Y_{.st}}{J_t \xi(c_{st}) + I_t}.$$

where  $X_{.st} := \sum_j X_{jst}$  and  $Y_{.st} := \sum_i Y_{ist}$ . For the case  $c_{st}$ , the  $2(S \cdot T)$  parameters are readily found via

$$\hat{\lambda}_{st} = \frac{X_{.st}}{J_t} \quad \text{and} \quad \hat{\gamma}_{st} = \frac{Y_{.st}}{I_t}.$$

## Unobserved Counts

Probably need some words about why we might not observed the count data.

When the data  $X_{jst}$  are observed as a binary response, instead of count data representing the number of prey species  $s$  that the predator ate in time period  $t$ , we can still estimates the parameters of interest  $\boldsymbol{\lambda}, \boldsymbol{\gamma}$ . Because some information is observed, we can treat the counts as missing and use the EM algorithm to find the maximum likelihood estimates of the observed data likelihood.

We denote the binary response that the predator did in fact eat at least one prey species  $s$  in time period  $t$  by  $Z_{jst} = 1(X_{jst} > 0)$ . Now, the observed data  $Z_{jst} \stackrel{\text{iid}}{\sim} \text{Bern}(1 - \exp\{-\lambda_{st}\})$ . Using this we can find the complete data likelihood, by first noting that the conditional distribution of the unobserved data  $X_{jst}$  given  $Z_{jst}, Y_{ist}, \boldsymbol{\lambda}, \boldsymbol{\gamma}$  is a truncated Poisson distribution

$$f_{X|Y,Z,\boldsymbol{\lambda},\boldsymbol{\gamma}}(x_{jst}) = \frac{\exp\{-\lambda_{st}\} \lambda_{st}^{x_{jst}}}{(1 - \exp\{-\lambda_{st}\}) x_{jst}!} 1(x_{jst} \geq 0) \quad \text{where} \quad \mathbb{E}_{[X|Y,Z]} X_{jst} = \frac{\lambda_{st} \exp\{\lambda_{st}\}}{\exp\{\lambda_{st}\} - 1}.$$

From this conditional distribution we get the joint distribution of  $X_{jst}, Z_{jst}$

$$f_{X,Z|\boldsymbol{\lambda}}(x_{jst}, z_{jst}) = \begin{cases} \exp\{-\lambda_{st}\}, & x_{jst} = 0 \text{ and } Z_{jst} = 0 \\ \frac{\exp\{-\lambda_{st}\} \lambda_{st}^{x_{jst}}}{x_{jst}!}, & x_{jst} > 0 \text{ and } Z_{jst} = 1 \\ 0 & \text{otherwise} \end{cases}$$

Take the expectation of the complete data log-likelihood with respect to  $f_{X|Y,Z,\boldsymbol{\lambda},\boldsymbol{\gamma}}(x_{jst})$  to get

$$\begin{aligned}
\mathbb{E}l_{comp} &= \mathbb{E} \log f_{X,Z|\lambda}(X_{jst}, z_{jst}) + \log f_{Y|\gamma}(y_{ist}) \\
&= \sum_{s=1}^S \sum_{t=1}^T \sum_{j=1}^{J_t} \mathbb{E} \log f_{X,Z|\lambda}(X_{jst}, z_{jst}) + \sum_{s=1}^S \sum_{t=1}^T \sum_{i=1}^{I_t} \log f_{Y|\gamma}(y) \\
&= \sum_{s,t,j} (-\lambda_{st} + z_{jst} \log(\lambda_{st}) \mathbb{E}X_{jst}) + \sum_{s,t} (-I_t \gamma_{st} + Y_{st} \log I_t \gamma_{st}) + \text{const} \\
&= \sum_{s,t} (-J_t \lambda_{st} + z_{st} \log(\lambda_{st}) \mathbb{E}X_{jst}) + \sum_{s,t} (-I_t \gamma_{st} + Y_{st} \log I_t \gamma_{st}) + \text{const}.
\end{aligned}$$

The EM algorithm requires iteratively solving  $(\lambda^{(k+1)}, \gamma^{(k+1)}) = \arg \max_{\lambda, \gamma} \mathbb{E}l_{comp}$ . Though no analytic solution to this maximization exists, one idea is to iteratively solve partial derivatives of  $\mathbb{E}l_{comp}$  set equal to zero until convergence. In fact, as we only need find parameter values that increase the observed likelihood, we forgo fully iterating to find the maximized values of  $(\lambda, \gamma)$  and instead perform just one step uphill within each EM iteration. This strategy is significantly less computationally intensive, thus generating a much faster generalized EM (GEM) algorithm.

The updating steps of the EM algorithm simply replace any occurrence of  $X_{st}$  with  $z_{st} \mathbb{E}X_{jst}$ . For instance, when maximizing under the hypothesis  $c_{st}$  we get  $\hat{\lambda}_{st} = z_{st} \mathbb{E}l_{comp} / J_t$ . In the case of the hypothesis  $c_t$ , where  $\lambda_{st} = c_t \gamma_{st}, \forall t$ , we iterate the following functions until convergence,

$$\hat{\gamma}_{st} = \frac{z_{st} \mathbb{E}X_{jst} + Y_{st}}{c J_t + I_t} \quad \text{and} \quad \hat{c} = \frac{\sum_t z_{st} \mathbb{E}X_{jst}}{J_t \sum_s \gamma_{st}}$$

updating the value of  $\lambda_{st}$  before each next step.

This GEM algorithm works well when values  $\lambda_{st}$  are relatively small and zeros are common in the data  $Z_{jst}$ . In this case, not too much information is lost since estimation of  $\mathbb{E}Z_{jst}$  can be estimated well by the proportion of observed zeros. On the other hand, if the predator consistently eats a given prey species, few to no zeros will show up in the observed data and  $\mathbb{E}Z_{jst}$  is estimated to be nearly 1. The loss of information is best seen by attempting to solve for  $\lambda_{st}$  in the equation  $1 = \mathbb{E}Z_{jst} = 1 - \exp\{-\lambda_{st}\}$ ; essentially  $\lambda_{st}$  is sent off to  $+\infty$ .

## Testing

All hypotheses are evaluated via a likelihood ratio test (LRT), with statistic

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

where  $\theta_0, \theta_1$  represent the parameters estimated under the null and alternative hypotheses, respectively. It is well known that the asymptotic distribution of  $\Lambda$  is a  $\chi_\rho^2$  distribution with  $\rho$  degrees of freedom. Under the EM algorithm we use  $L_{observed}(Z, Y)$  as the likelihood in the calculation of  $\Lambda$ .

The degrees of freedom  $\rho$  are set equal to the number of free parameters available in the stated hypotheses under question. If we put the null hypothesis to be  $H_0 : \lambda_t = c_t \gamma_t, \forall t$  and

contrast this against  $H_1 : \lambda_{st} = c_{st}\gamma_{st}$  then there are  $\rho = 2(S \cdot T) - S \cdot T - T = S \cdot T - T$  degrees of freedom. The alternative hypothesis previously stated is not in fact fit by estimating  $c_{st}$ , but is instead fit as if  $c_{st}\gamma_{st}$  were  $S \cdot T$  unique parameters independent of  $\lambda_{st}$  for all  $s, t$ , thus providing  $2(S \cdot T)$  total parameters for the hypothesis we denote by  $c_{st}$ .

A set of hypotheses is determined by the p-value of the  $\chi^2_\rho$  distribution. We choose the common value of  $\alpha = 0.05$  as our level of significance. Hence, if  $\mathbb{P}(\chi^2_\rho > \Lambda) < 0.05$  the null hypothesis is rejected in favor of the alternative hypothesis.

## Simulations

Our simulations hypothesize two prey species, and five time points. Of the hierarchy of hypotheses, we simulate data under three null hypotheses:  $\{c_s, c_t\}$ . Sample sizes are chosen randomly from four overlapping levels. Let “small” sample sizes be randomly sampled numbers in  $[20, 50]$ , “medium” encompass  $[30, 75]$ , “large”  $[50, 150]$ , and “huge”  $[100, 200]$ . Hence, we randomly sample prey and predator gut count observations for each time period from one of the sample size levels, then cycle through all hypotheses. This is repeated for each level of sample size. We simulate data for each of the 8 scenarios above under both the standard maximum likelihood theory and under the EM algorithm. Only the details of the EM algorithm are presented here, as the other simulations only reinforce the theory of point estimation; that our estimators are uniformly minimum-variance and unbiased (UMVUE).

For all simulated data, the true parameter values for the rate at which prey species are encountered in the wild are fixed to be  $\gamma_{st} := \pi, \forall s, t$ . The values of  $\lambda_{st}$  are set for each null hypothesis. Under the hypothesis  $c_s$ , the ratio of rates varies by species  $s$  only, so we put  $\lambda_{1t} := \sqrt{2}$  and  $\lambda_{2t} := \pi$ . Hence,  $c_1 := \sqrt{2}/\pi \approx 0.45$  and  $c_2 := 1$ . For the second hypothesis, the ratio of rates varies by time  $t$ . Here, we put  $\lambda_{st} := t$  for  $t \in \{1, \dots, 5\}$ .

As noted above we find that the EM algorithm performs well when the parameter values of  $\lambda_{st} = \xi(c_{st})\gamma_{st}$  are small. Figure 3 contains density plots of the EM algorithm’s estimates for the values  $c_s, c_t$  for small and huge sample sizes, respectively. When data are simulated under the null hypothesis  $c_s$ , we find, even for the small sample size that point estimates of  $c_s$  are quite accurate. Though, when the values of  $\lambda_{st}$  are large enough to make 0s in the simulated data less common, the algorithm occasionally over-estimates the true values of  $c_t$  despite the increased sample size. The plots of  $\gamma_{st}$  under the EM algorithm are not given as we do not consider missing data in the estimation of these parameters, hence they are UMVUE as before.

## Conclusion

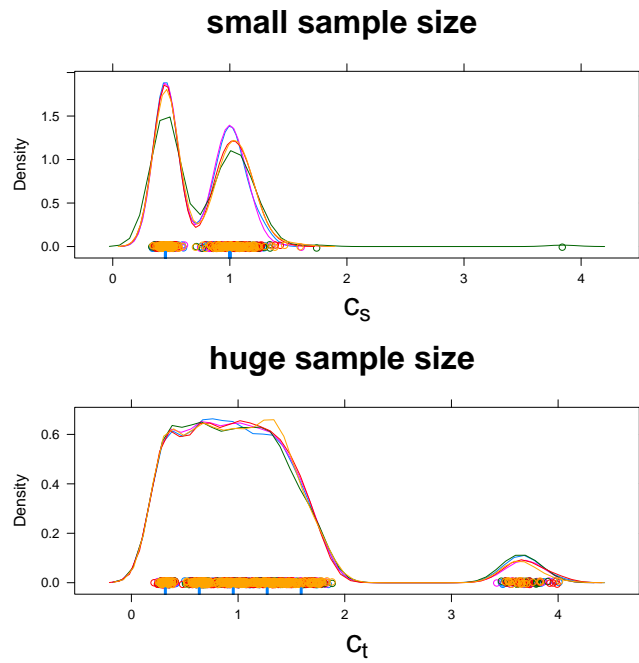


Figure 3: Density plots of the EM algorithm's estimates of  $c_s$  and  $c_t$  with small and huge sample sizes, respectively. The estimates near 4 for both of the plots represent samples for which few zeros appeared in the data and the EM algorithm had a particularly difficult time estimating the true parameter value.