Size-dependent death rate of the Toadis-uglibus frog

Jonathan Eagle, Matthew Barrett, Pierre Poupart-Lafarge

May 4, 2021

1 Introduction

Frogs of the specie toadis uglibus are currently endangered, and thus conservationists are looking for a method to ensure their survival. Frogs are hunted by predators such as small mammals, snakes and birds. It is believed that there is a critical period during which the frogs are most vulnerable to such predators, when they are too large to hide, but still too small to deter their predators. Due to financial limits on the project, the optimal method to aid the species survival is to take the frogs into captivity at their most vulnerable and release them when they are less so. Much like the use of canaries by miners to detect poisonous gas in tunnels, frog's amphibious nature, which makes them sensitive to changes in the environment, means they function as a primary indicator for the current health of the environment thus enabling immediate action. Also holding a vital role in the food chain as both predator and prey, their survival is key. An experiment was carried out in a large, enclosed pond assessing the effect of the size of the frog on the probability of it being killed. Using the collected data, we first fit 2 different models using maximum likelihood estimation (MLE), assessed their fits both visually and analytically and finally used our models to interpret the critical size as discussed, along with the predictive uncertainty.

2 Model and method

2.1 Distributional assumptions

A total of i experiments were carried out in which 10 frogs of length s_i were placed in a large, enclosed pond, and were checked up on 3 days later. The number of frogs killed in each experiment was recorded and used in analysis. A binomial distribution counts the number of successes from n independent trials with probability p of success in each trial. Therefore, the number of deaths from the 10 frogs in experiment i can be assumed to be distributed such that

$$\text{kill}_i \sim \text{binom}(p_i, 10)$$

where $p_i = P(s_i)$ is the probability of a frog being killed given its size, s_i. It is important to note that under this model we are assuming that size is the only factor that will affect whether a frog is killed (in reality colour, location and whether a predator has just eaten a frog may also play an important role).

2.2 The modified logistic function

The logistic model has long been used as a base method of inferring population dynamics in a closed system. Modifications have been made to the original logistic model due to the change in thing we are measuring (kills not population) and an increase in parameters to ensure the most effective fit. Our first model of the probability of a frog being killed in terms of its size is therefore given by a modified logistic function

$$P_L(s) = \frac{e^{\varepsilon(\phi-s)}}{1 + e^{\beta\varepsilon(\phi-s)}}$$

where s is the size of the frog. Also, $\phi > 0$ is a location parameter such that $P(s = \phi) = 0.5$, $\varepsilon < 0$ controls the rate of change of $P(s)$ with s and $\beta > 1$ controls the asymmetry of the function.
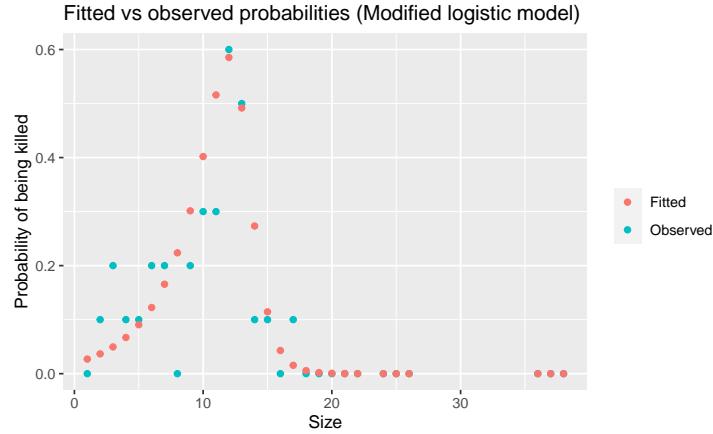
2.3 Ricker model

Alternatively we have the generalised Ricker model as a means of estimating the probability P(s):

$$P_R(s) = b \left\{ \frac{s}{a} \exp\left(1 - \frac{s}{a}\right) \right\}^{\alpha}$$

where $a > 0$, $b > 0$ and $\alpha > 0$ are parameters to be estimated. Developed originally by Dr Bill Ricker in 1954 as a means of understanding stock and recruitment in fisheries, the ricker model has since become a backbone in discrete population modelling in a closed environment. Appearing in its raw form as a difference equation, stability and equilibria can be analysed more easily. In our case this is the value of interest; the stationary point where the frogs are too small to protect themselves but too big to hide from predators. We will investigate this analytically using the maximum likelihood estimator.

2.4 Maximum likelihood estimation

Having specified 2 different ways of calculating the probability of a frog of a given size being killed, we can find the log likelihoods given by the 2 models. We aim to find the maximum likelihood estimates (MLEs) of the parameters of each model using the BFGS method and assess which is more effective at fitting the data. BFGS is an iterative method similar to Newton's method but differs in that it only approximates the hessian matrix rather than calculating it directly. Using this method we minimise the negative log likelihoods; the equivalent of maximising the log likelihoods. Under the modified logistic model we find the MLEs to be $(\phi, \varepsilon, \beta) = (12.955, -0.302, 4.408)$, and we can visualise the fitted probabilities given by these MLEs against the true proportions of frogs that were killed

Under the generalised Ricker model we find the MLEs to be $(a, b, \alpha) = (8.092, 0.286, 3.663)$, and again visualising the fitted and observed probabilities gives

2
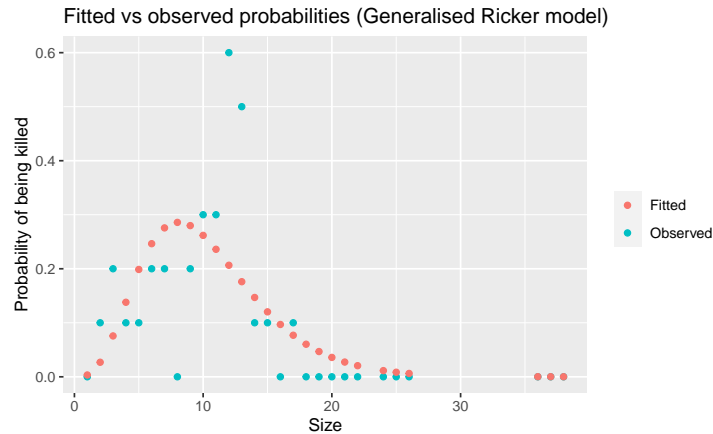
2.5   Model checking

Visually it appears that the modified logistic model is a much better fit for our raw data. The extreme observations at sizes 12 and 13 seem to be discarded by the Ricker model as outliers, which is especially problematic for our purposes since the maximum is what we are ultimately interested in. Furthermore, the sharp cut off then displayed by the modified logistic model following the peak fits the raw data which has a 0-kill toll from 18cm onwards. Hence, the modified logistic model seems to fit the data better than the Ricker model.

To confirm what we are seeing in the plots we can calculate the Akaike information criterion (AIC) for each of the models, giving us a means of comparing the relative quality of the 2 models. We find that the AIC for the modified logistic model is 55.4 and the AIC for the Ricker model is 68.9. The AIC is lower for the modified logistic model which provides further evidence that this model fits the data better.

3   Budget allocation

3.1   Size of maximal predation

Due to the limited budget the conservationists have, we seek to find the frogs which are most vulnerable to being eaten by predators to allow to focus the limited resources on protecting these frogs. In particular we want to compute the size for which the probability of the frog being killed is highest i.e. we want to maximise

$$P_L(s) = \frac{e^{\varepsilon(\phi - s)}}{1 + e^{\beta\varepsilon(\phi - s)}}$$

with respect to s. Equivalently, we will maximise

$$\log(P_L(s)) = \log\left(\frac{e^{\varepsilon(\phi - s)}}{1 + e^{\beta\varepsilon(\phi - s)}}\right) = \varepsilon(\phi - s) - \log(1 + e^{\beta\varepsilon(\phi - s)}).$$

Differentiating with respect to s and setting this equal to 0 we obtain

$$-\varepsilon - \frac{-\beta\varepsilon e^{\beta\varepsilon(\phi - s)}}{1 + e^{\beta\varepsilon(\phi - s)}} = 0 \iff -1 - \frac{-\beta e^{\beta\varepsilon(\phi - s)}}{1 + e^{\beta\varepsilon(\phi - s)}} = 0 \iff \beta e^{\beta\varepsilon(\phi - s)} = 1 + e^{\beta\varepsilon(\phi - s)}$$

$$\iff (\beta - 1)e^{\beta\varepsilon(\phi - s)} = 1$$

$$\iff \beta\varepsilon(\phi - s) = -\log(\beta - 1)$$

$$\iff s = \phi + \frac{\log(\beta - 1)}{\beta\varepsilon}.$$

By plugging our MLEs into this equation we get an estimate of the size of maximal predation $\hat{s}^* = 12.03$.

3.2   Uncertainty

To give the conservationists more information on how much they can trust our estimate, we want to provide them with the distribution of $\hat{s}^*$. As the asymptotic distribution of $\hat{s}^*$ cannot easily be derived analytically, we will study it by sampling m times from an estimate of the asymptotic distribution of the MLE, $\hat{\theta}_c$. From this we can compute $\hat{s}^*$ and study its distribution.

To reduce the uncertainty that arises from sampling we take a large $(m = 10^6)$ number of samples. We began by taking sample vectors $\hat{\theta}_c^i$ from the approximate asymptotic distribution of $\hat{\theta}_c$, which is a multivariate Gaussian centered at the MLE and with covariance matrix given by the inverse of the Hessian of the negative log likelihood evaluated at the MLE. We then transformed the vectors $\hat{\theta}_c^i$ back to the original parameters $\hat{\theta}^i$ i.e. with $\hat{\theta}_c^i = \left(\hat{\phi}_c^i, \hat{\epsilon}_c^i, \hat{\beta}_c^i\right)$ we have

$$\hat{\theta}^i = \left(\exp\left(\hat{\phi}_c^i\right), -\exp\left(\hat{\epsilon}_c^i\right), \exp\left(\hat{\beta}_c^i\right) + 1\right)$$

For each vector $\hat{\theta}^i$ we were able to compute $\hat{s}_i^*$ using the following function,

$$\hat{s}_i^* = g\left(\hat{\theta}^i\right) = g\left(\hat{\phi}_i, \hat{\epsilon}_i, \hat{\beta}_i\right) = \frac{log(\hat{\beta}_i - 1)}{\hat{\beta}_i \hat{\epsilon}_i} + \hat{\phi}_i$$

Using all of the values of $\hat{s}_i^*$ we can visualise the approximate distribution of $\hat{s}^*$
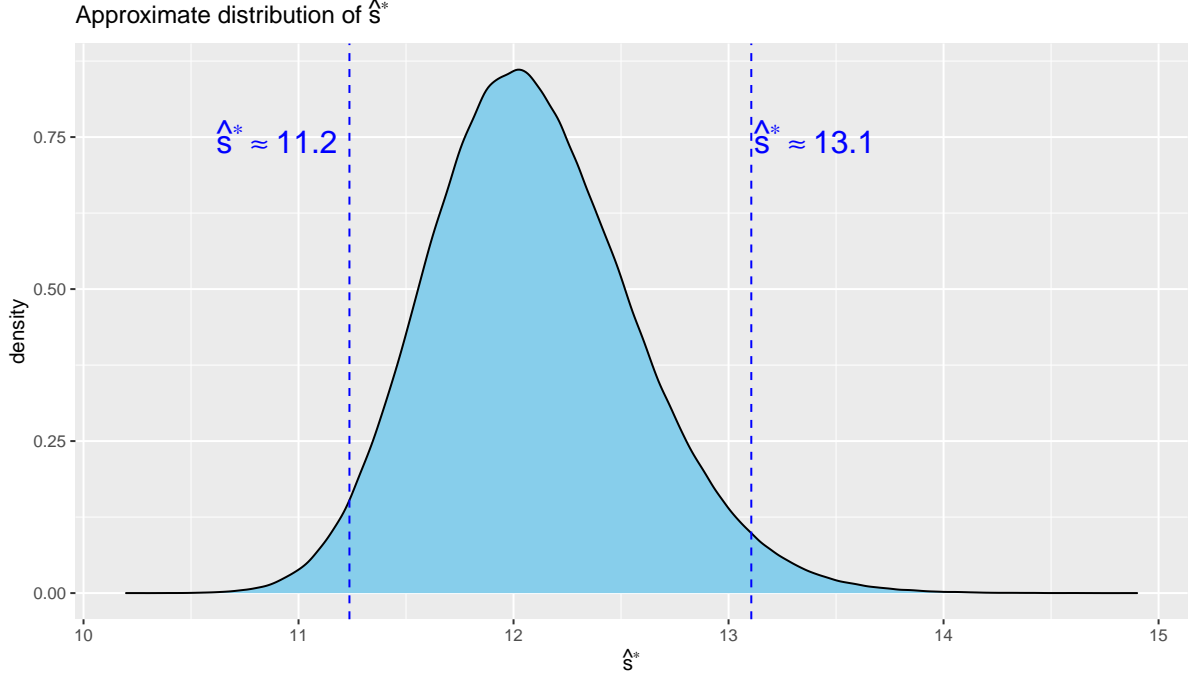
We also computed the standard deviation (0.48) and the plot shows the 95% confidence interval for $\hat{s}^*$ is (11.2, 13.1).

3.3 Advice

In the last section we showed that we can be 95% certain that the most vulnerable frogs are between 11.2cm and 13.1cm in length. Therefore our recommendation is that all frogs within this range should be held in captivity until they are greater than 13.1cm. However, without more specific information about the budget we can't be sure whether the conservationists have enough resources to protect all frogs within this range, or whether they have sufficient resources to protect all frogs in an even greater range. To account for this we show the intervals relating to varying levels of confidence:

Level of confidence	Interval	Level of confidence	Interval
5%	(12.03, 12.09)	55%	(11.72, 12.43)
10%	(12.00, 12.12)	60%	(11.69, 12.48)
15%	(11.97, 12.15)	65%	(11.65, 12.53)
20%	(11.94, 12.18)	70%	(11.60, 12.58)
25%	(11.91, 12.21)	75%	(11.56, 12.64)
30%	(11.88, 12.25)	80%	(11.50, 12.71)
35%	(11.85,12.28)	85%	(11.44, 12.80)
40%	(11.82, 12.32)	90%	(11.36, 12.92)
45%	(11.79, 12.35)	95%	(11.24, 13.11)
50%	(11.76, 12.39)	99%	(11.00, 13.49)

Hence we advise the conservationists to choose the largest of these intervals such that their budget allows them to protect all frogs in the given range. More specific intervals can be calculated upon request to ensure that the entire budget is used effectively.

4 Appendix

```r
---
title: "inference project"
output: html_document
---

```{r}
load(file = url("https://mfasiolo.github.io/TOI/frogs.RData")) #loads our dataset "
    frogs"

# function  to calculate probabilities of different sized frogs being killed under a
    modified logistic model
p.L<- function (phi, eps, beta) {
  s<- frogs[,1]  #vector containing the different sizes of the frogs
  return (c(exp(eps*(phi-s))/(1 + exp(beta*eps*(phi-s))))) # return s vector of
    probabilities
}

# function  to evaluate the negative log likelihood under the modified logistic model
log.lik.L<- function (x) {
  phi<- exp(x[1])  #appropriate transformations ensure that phi > 0,
  eps<- -exp(x[2])  #epsilon < 0
  beta<- exp(x[3]) + 1  #beta > 1
  p<- p.L(phi, eps, beta) #calculates the probabilities of the different sized
    frogs being killed
  obs<- frogs[,2]  #our observations of the number of frogs killed in each
    experiment
  a<- sum(dbinom(obs, size = 10, prob = p, log=TRUE)) #computes the log likelihood
  return (-a)  # return s negative log likelihood
}

logist<- optim(c(2.5,-1.4,2.5), log.lik.L, method = "BFGS", hessian = TRUE) #finds
    maximum likelihood estimates of the modified logistic model parameters
p.L.hat<- p.L( exp(logist$par[1]), -exp(logist$par[2]), exp(logist$par[3])+1 ) #
    computes the fitted probabilities with these MLEs

library (ggplot2) #package to enable better looking visualisations
df<- data.frame(Size=c(frogs[,1], frogs[,1]), Prob=c(frogs[,2]/10, p.L.hat), grp=
    rep(c("Observed", "Fitted"), each=28 ))
ggplot(df, aes(x=Size, y=Prob, group=grp)) +
  geom_point(aes(color=grp)) +
  ggtitle("Fitted vs observed probabilities (Modified logistic model)") +
  ylab("Probability of being killed") +
  theme(legend.title = element_blank())
#we visualise the actual proportions of frogs killed in blue vs the fitted
    probabilities in red
```

```{r}
# function  to calculate probabilities of different sized frogs being killed under a
    generalised Ricker model
p.R<- function (a, b, alpha) {
  s<- frogs[,1]  #vector containing the different sizes of the frogs
  return (c(b*( (s/a)*exp(1-s/a) )^alpha)) # return s vector of probabilities
}

# function  to evaluate the negative log likelihood under the generalised Ricker
    model
log.lik.R<- function (x) {
  a<- exp(x[1])  #transformations to ensure: a > 0
  b<- exp(x[2])  #b > 0
  alpha<- exp(x[3])  #alpha > 0
  p<- p.R(a, b, alpha) #calculates the probabilities of the different sized frogs
    being killed
  obs<- frogs[,2]  #our observations of the number of frogs killed in each
    experiment
  a<- sum(dbinom(obs, size = 10, prob = p, log=TRUE)) #computes the log likelihood
  return (-a)  # return s the negative log likelihood
}

ricker<- optim(c(2,-1.4,1.1), log.lik.R, method = "BFGS", hessian = TRUE) #finds
    maximum likelihood estimates of the generalised Ricker model parameters
```

```r
58  p.R.hat<- p.R(exp(ricker$par[1]), exp(ricker$par[2]), exp(ricker$par[3])) #computes
        the fitted probabilities with the MLEs
59
60  library(ggplot2) #package to enable better looking visualisations
61  df<- data.frame(Size=c(frogs[,1], frogs[,1]), Prob=c(frogs[,2]/10, p.R.hat), grp=
        rep(c("Observed", "Fitted"), each=28 ))
62  ggplot(df, aes(x=Size, y=Prob, group=grp)) +
63    geom_point(aes(color=grp)) +
64    ggtitle("Fitted vs observed probabilities (Generalised Ricker model)") +
65    ylab("Probability of being killed") +
66    theme(legend.title = element_blank())
67  #we visualise the actual proportions of frogs killed in blue vs the fitted
        probabilities in red
68
69
70  2*length(logist$par) + 2*log.lik.L(logist$par) #AIC for logistic model
71  2*length(ricker$par) + 2*log.lik.R(ricker$par) #AIC for ricker model
72
73  ```
74
75
76  ```{r}
77  # function  to calculate the size of maximal predation
78  g<-  function (x) {
79    s.hat.star<- log(x[3]-1)/(x[3]*x[2]) + x[1]
80     return (s.hat.star)
81  }
82  g(c(exp(logist$par[1]), -exp(logist$par[2]), exp(logist$par[3])+1))   #estimate of
        size where predation is maximal
83
84  covar<- solve(logist$hessian) #approximate asymptotic covariance with the inverse
        of the hessian of the negative log likelihood evaluated at MLE
85  mean<- logist$par #approximate the asymptotic mean with MLE
86
87  library(MASS)     #provides function  to simulate multivariate normal
88  samples<- mvrnorm(1000000, mean, covar)   #1.sample from approximate distribution
89  theta.hat<- cbind(exp(samples[,1]),-exp(samples[,2]),exp(samples[,3])+1)  #2.
        convert all samples to original parameters
90  s.i<- apply(theta.hat, 1, g)  #3.compute s.hat_i for each sample
91
92  sd(s.i)  #compute standard deviation of samples
93  ci<- quantile(s.i, probs = c(0.025,0.975))  #95% confidence interval
94
95  #visualising distribution of s.hat^* with 95% CI bounds
96  library(ggplot2)
97  library(latex2exp)
98  df<- data.frame(shat=s.i)
99  ggplot(df, aes(x=shat)) +
100   geom_density(fill="skyblue") +
101   geom_vline(aes(xintercept=ci[1]), linetype="dashed", color="blue") +
102   geom_vline(aes(xintercept=ci[2]), linetype="dashed", color="blue") +
103   xlab(TeX("$\\hat{s}^*$")) +
104   ggtitle(TeX("Approximate distribution of $\\hat{s}^*$")) +
105   annotate(geom = "text", x=10.9, y=0.75, size=6, col="blue", label=TeX("$\\hat{s}^
        *\\approx 11.2$")) +
106   annotate(geom = "text", x=13.4, y=0.75, size=6, col="blue", label=TeX("$\\hat{s}^
        *\\approx 13.1$"))
107
108 #finding intervals for varying levels of confidence to allow conservationists to
        adapt which frogs they protect based on their budget
109 intervals<- matrix(0, nrow=20, ncol=2)
110 for (i in 1:19) {
111   intervals[i,]<- quantile(s.i, probs = c( (50-5*i/2)/100, (50+5*i/2)/100 )) #CIs
        5%, 10%,.., 95%
112 }
113 intervals[20,]<- quantile(s.i, probs = c( (50-99/2)/100, (50+99/2)/100 )) #99% CI
114 intervals
115 ```
```