

Problem Set 3b

Model Selection by Maximum Likelihood

We've been given a long time-series of data regarding a population's dynamics and are tasked with determining whether the population exhibited density-dependence and, if so, of what form it is. In particular, we need to evaluate the relative support for three models of population growth:

1. density-independent growth (as in PS #3a),
2. density-dependent growth as described by the classic Ricker model, and
3. density-dependent growth as described by the so-called theta-logistic Ricker model.

We will assume that observation error is negligible such that variance not attributable to density is primarily due to process error. The data we will use are from population abundance surveys of great tits (*Parous major*) in the Netherlands (*greattit.csv*, posted to our course website). You might want to review your readings (specifically Morris and Doak pg. 108-118) prior to working on this.

Part A

Load the time series:

```
dat <- read.csv('greattit.csv', header = TRUE)
Year <- dat[, 1]
N_obs <- dat[, 2]
```

The first column is survey year, the second column is the population count (N). Make a plot of N versus $Year$ to get an idea of what the raw data look like. Create a new variable (e.g., `log_lambda_obs`) to store values of $\log(\lambda t)$, and then calculate this value at each year of the time series, i.e. $\log(\lambda t) = \log(N_{t+1}/N_t)$. Make a scatter-plot of $\log(\lambda t)$ versus N_t to get a visual idea of how annual growth rate varies with population size.

Part B

Create three new functions (see hint section of PS#3a for tips on making functions), calling them something like `m1`, `m2` and `m3`. These functions will correspond to our three models of population growth (density-independent, Ricker, and theta-logistic), and will be used to estimate $\log(\lambda t)$ as a function of N_t . You know the equations for calculating N_{t+1} as a function of N_t and the parameters r , K and θ (see below), so you will need to rearrange these equations to solve for $\log(\lambda t)$. Each of your functions should have the argument of `b` which will contain the model-specific vector of the appropriate parameters r , K and θ , and the independent variable x (which corresponds to N_t). So, for example, the first function for density-independent growth will look like:

```
m1<-function(x){ bm1 + 0 * x }
```

where `bm1` will contain only a single parameter corresponding to r .

Part C

Use nonlinear least squares (`nls()`) to estimate the parameters for each of the three models (see below for hints on how). Also calculate the associated standard deviation (σ) for each model, estimated as the square root of the mean squared residuals (use `resid(fit)` to extract the residuals of your fitted model). Do this for each of the three models. Make a figure with 3 panels (or subplots), each showing a scatter plot of $\log(\lambda t)$ versus N_t and with the least-squares estimate of the expected value from each function plotted as a line (see hints below). Which one looks like the “best fit”?

Part D

Given the least-squares parameter estimates for each of the models, calculate the associated negative log-likelihood values. To do this you will need the equation giving the log-likelihood equation for a normal random variable (see hints below), your predicted and observed values of $y = \log(\lambda t)$ at each N_t , and the standard deviation (σ) associated with each model fit. Then calculate the AIC, ΔAIC and AIC weights for each model. Concatenate the AIC results from the three models to a single matrix. Which of the three models has the most support?

Hopefully-helpful code and equations

Code for nonlinear least squares:

```
m1.fit <- nls(y ~ b0 + 0 * x, start=list(b0 = 0))
b.m1 <- coef(m1.fit)
sigma.m1 <- sqrt(mean(resid(m1.fit)^2))
```

Code to plot data and fitted model:

```
plot( x, y, xlab = 'N_t', ylab = 'log(N_t + 1 / N_t)' )
m1 <- function(x){ b.m1 + 0 * x}
curve(m1, c( min(x):max(x) ) )
```

Density-independent model	$N_{t+1} = N_t e^r$
Density-dependent Ricker model	$N_{t+1} = N_t e^{r\left(1 - \frac{N_t}{K}\right)}$
Density-dependent Theta-logistic	$N_{t+1} = N_t e^{r\left(1 - \left(\frac{N_t}{K}\right)^\theta\right)}$

Negative log-likelihood for Normal random variable (q is the # of observations, P_t is the predicted value at time t)	$-\ln(\mathcal{L}) = \frac{q}{2} \ln(2\pi\sigma^2) + \frac{1}{2\sigma^2} \sum_{t=1}^q \left(\ln\left(\frac{N_{t+1}}{N_t}\right) - P_t \right)^2$
AIC (for a model with p estimated parameters)	$2p - 2\ln(\mathcal{L})$
ΔAIC (for model m)	$\text{AIC}_m - \min(\text{AIC})$
AIC weight (for model m)	$\frac{e^{-\frac{1}{2}\Delta\text{AIC}_m}}{\sum_i e^{-\frac{1}{2}\Delta\text{AIC}_i}}$