Biomass model and likelihood

Assessment workshop

Arni Magnusson

Hafro, 16–24 Jan 2012

Outline

1. Biomass model I - dynamics, data, least squares, uncertainty
2. Likelihood - relative probability, support, combine data sources
3. Estimation - MLE, log likelihood, confidence interval
4. Biomass model II - normal distribution, likelihood inference

Outline

1. Biomass model I - dynamics, data, least squares, uncertainty
2. Likelihood - relative probability, support, combine data sources
3. Estimation - MLE, log likelihood, confidence interval
4. Biomass model II - normal distribution, likelihood inference

Biomass dynamics

$$B_{t+1} \quad = \quad B_t \quad + \quad g(B_t) \qquad - \quad C_t$$

Surplus production can also be of interest in age-structured models

Schaefer (1954):

$$B_{t+1} \quad = \quad B_t \quad + \quad r\,B_t\left(1-\frac{B_t}{K}\right) \quad - \quad C_t$$

Parameters and reference points

B_{init} initial biomass

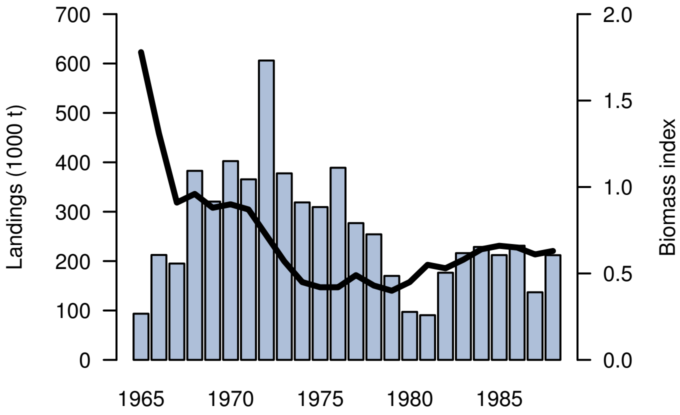
r maximum growth rate

K carrying capacity

q catchability coefficient

$$
\begin{aligned}
B_{\text{MSY}} &= K/2 \\
MSY &= rK/4 \\
u_{\text{MSY}} &= r/2
\end{aligned}
$$

Namibian hake



Fit model in spreadsheet

1. Get **data**
   . . . and plot

   $t \quad B \quad C \quad I \quad Ifit \quad res2$

2. Define **parameters**
   . . . and consider transforming

   $B_{\text{init}} \quad r \quad K \quad q$

3. Calculate **predictions**
   . . . and fit by eye

$$B_{t+1} = B_t + rB_t\left(1-\frac{B_t}{K}\right) - C_t$$
$$\hat{I}_t = qB_t$$

4. Calculate **RSS**
   . . . and optimize

$$RSS = \sum\left(\log I_t - \log \hat{I}_t\right)^2$$

Uncertainty

We have point estimates, excellent!

But what about **uncertainty**?

Outline

1. Biomass model I - dynamics, data, least squares, uncertainty
2. Likelihood - relative probability, support, combine data sources
3. Estimation - MLE, log likelihood, confidence interval
4. Biomass model II - normal distribution, likelihood inference

Likelihood concepts

Relative probability

$$P(A_1) = 0.5 \qquad P(A_2) = 0.3 \qquad P(A_3) = 0.2$$

$$L(A_1) = 500 \qquad L(A_2) = 300 \qquad L(A_3) = 200$$

$$L(A_1) = 0.005 \qquad L(A_2) = 0.003 \qquad L(A_3) = 0.002$$

Likelihood concepts

Expresses how well the data **support** some parameter value or hypothesis

$$L\left(\theta \middle| \mathrm{data}\right)$$

Like *RSS* but even more useful:

not only point estimate, but also **uncertainty**

Likelihood concepts

We can fit a model to many types of data at once and **combine** the likelihood components with simple multiplication

$$L = L_1 \times L_2 \times \cdots$$

Unified framework, for simple or complex models

Likelihood concepts

Choose between models with different number of parameters

$$2 \log \frac{L_1}{L_0} \quad \sim \quad \chi^2_{\Delta df}$$

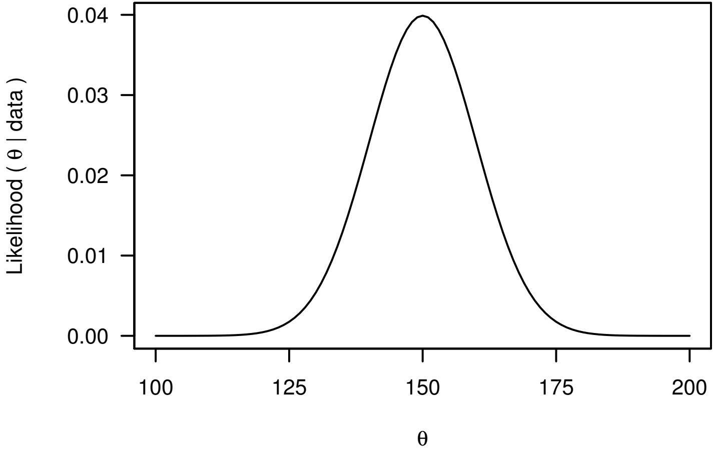
$$\text{AIC} \quad = \quad -2 \log L \; + \; 2k$$

$$\text{BIC} \quad = \quad -2 \log L \; + \; \log(n) k$$

Outline

1. Biomass model I - dynamics, data, least squares, uncertainty
2. Likelihood - relative probability, support, combine data sources
3. Estimation - MLE, log likelihood, confidence interval
4. Biomass model II - normal distribution, likelihood inference

Maximum likelihood estimation



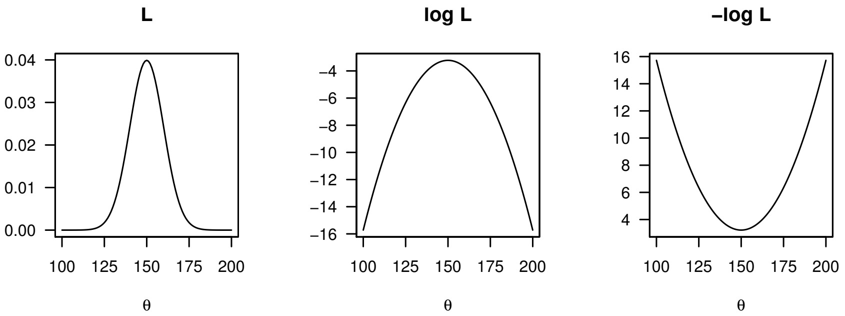
Log likelihood

Log transformation makes things easier

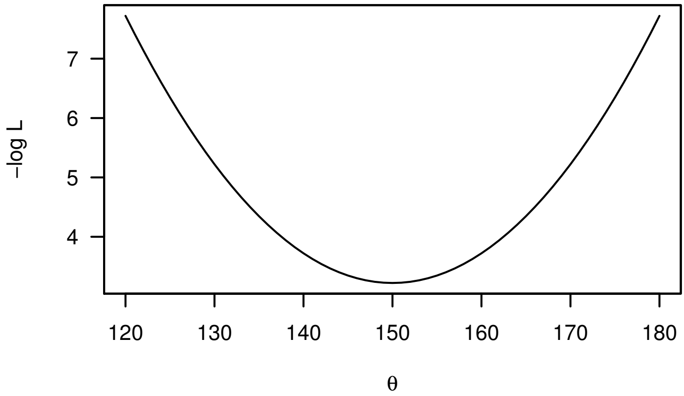
$$
\begin{aligned}
L(\theta|\text{data}) &= p(\text{data}|\theta) \\
& p(y_1, \ldots, y_n|\theta) \\
& p(y_1|\theta) \times \cdots \times p(y_n|\theta) \\
& \prod p(y_i|\theta)
\end{aligned}
$$

$$
\log L(\theta|\text{data}) = \sum \log p(y_i|\theta)
$$

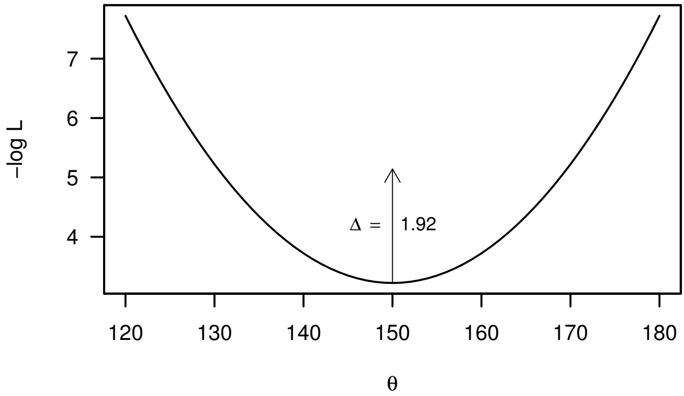
Log likelihood



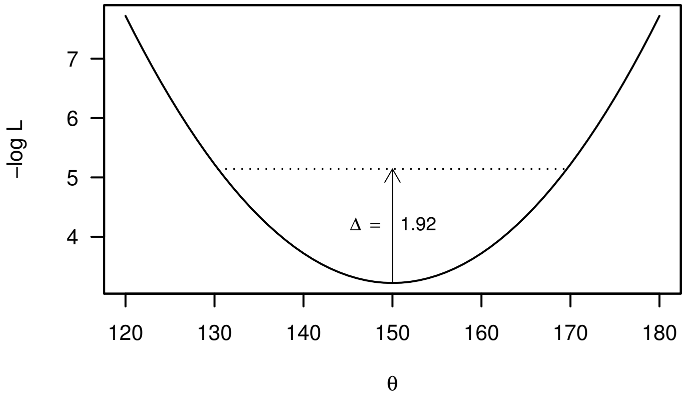
Confidence interval



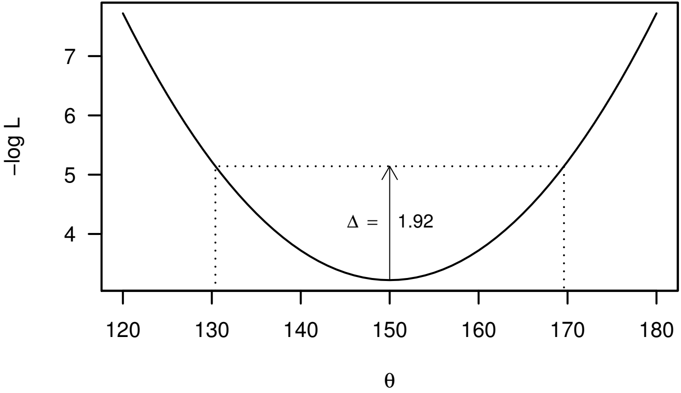
Confidence interval



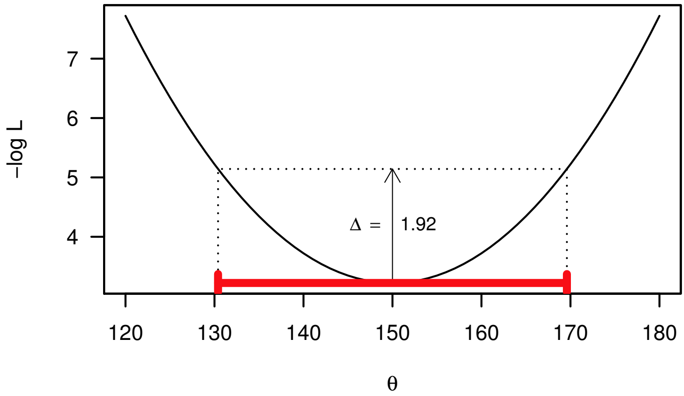
Confidence interval



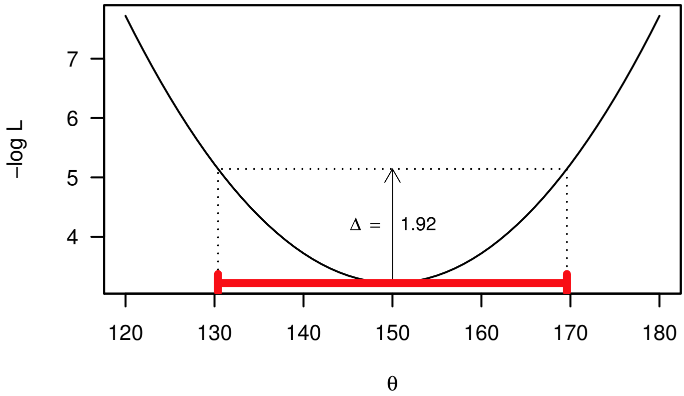
Confidence interval



Confidence interval



Confidence interval



$0.5\chi^2_{df=1} = 1.92$ for 95% confidence interval

Outline

1. Biomass model I - dynamics, data, least squares, uncertainty
2. Likelihood - relative probability, support, combine data sources
3. Estimation - MLE, log likelihood, confidence interval
4. Biomass model II - normal distribution, likelihood inference

Normal distribution

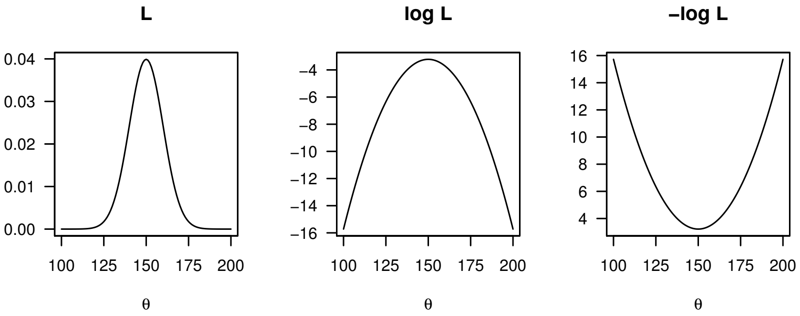
$$
\begin{aligned}
f(x) &= \frac{1}{\sigma\sqrt{2\pi}}\; e^{-\frac{(x-\mu)^2}{2\sigma^2}} \\
p(y_i|\theta) &= \frac{1}{\sigma\sqrt{2\pi}}\; e^{-\frac{(y_i-\mu_i)^2}{2\sigma^2}} \\
L(\theta|y) &= \prod\left(\frac{1}{\sigma\sqrt{2\pi}}\; e^{-\frac{(y_i-\mu_i)^2}{2\sigma^2}}\right) \\
-\log L &= \left[0.5n\log(2\pi)\right] \quad + \quad n\log\sigma \quad + \quad \frac{\sum(y_i-\mu_i)^2}{2\sigma^2} \\
&= \left[0.5n\log(2\pi)\right] \quad + \quad n\log\sigma \quad + \quad \frac{RSS}{2\sigma^2}
\end{aligned}
$$

dnorm in R

```
L <- prod(dnorm(y, mu, sigma))
```

```
neglogL <- -sum(dnorm(y, mu, sigma, log=TRUE))
```

dnorm in R




```
dnorm(theta,
m=150, s=10)
```

```
dnorm(theta,
m=150, s=10, log=T)
```

```
-dnorm(theta,
 m=150, s=10, log=T)
```

Back to our Schaefer model

Use likelihood inference to evaluate uncertainty

Let's calculate 95% confidence interval for K

Fit model in spreadsheet

1. Get **data**

   . . . and plot

   $t \quad B \quad C \quad I \quad Ifit \quad res2$

2. Define **parameters**

   . . . and consider transforming

   $B_{\text{init}} \quad r \quad K \quad q \quad \boldsymbol{\sigma}$

3. Calculate **predictions**

   . . . and fit by eye

$$B_{t+1} = B_t + rB_t\left(1 - \frac{B_t}{K}\right) - C_t$$
$$\hat{I}_t = qB_t$$

4. Calculate **neglogL**

   . . . and optimize

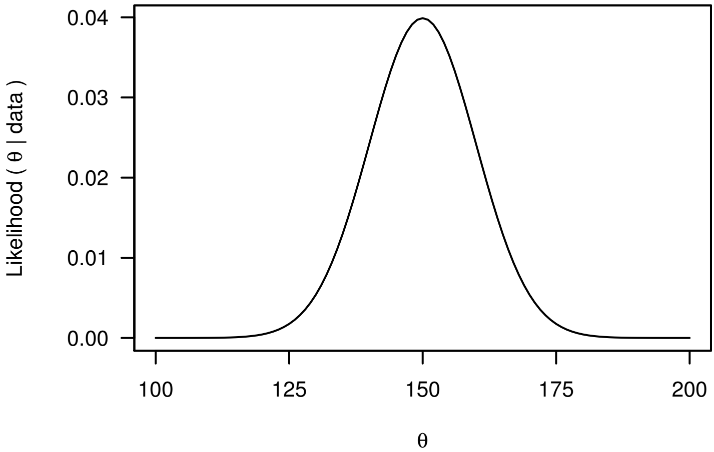
$$-\log L = 0.5n\log(2\pi) + n\log\sigma + \frac{RSS}{2\sigma^2}$$

Profile likelihood

1. Fix θ (a parameter of interest) at some value
2. Minimize $-\log L$ by estimating all other parameters
3. Save this value of $-\log L$

Repeat over a range of θ values

Profile likelihood



Discussion

- Difference between RSS and likelihood
- $B_{\text{init}} = K$?
- 95% confidence interval for MSY $rK/4$
- 95% confidence interval for B in final year

Discussion

- Difference between RSS and likelihood
- $B_{\text{init}} = K$?
- 95% confidence interval for MSY $\qquad rK/4$
- 95% confidence interval for B in final year

Discussion

- Difference between RSS and likelihood
- $B_{\text{init}} = K$?
- 95% confidence interval for MSY $\quad rK/4$
- 95% confidence interval for B in final year

Discussion

- Difference between RSS and likelihood
- $B_{\text{init}} = K$?
- 95% confidence interval for MSY $\quad rK/4$
- 95% confidence interval for B in final year