Notes on Likelihood functions for model fitting

Likelihood functions

For a parameter θ and data $x_1,...,x_n$ from an independent sample with a probability density function $f()$ the likelihood is given by

$$L(\theta) = \prod_{i=1}^{n} f(x_1,...,x_n \mid \theta)$$

And the log likelihood by

$$l(\theta) = \sum_{i=1}^{n} \ln f(x_1,...,x_n \mid \theta)$$

Normal distribution

If $x_1, x_2 ..., x_n$ is an independent sample from a normal distribution with mean μ and variance σ^2, then the joint distribution of $x_1, x_2 ..., x_n$ is

$$f(x_1,...,x_n \mid \mu,\sigma) = \prod_{i=1}^{n} \frac{1}{(2\pi)^{\frac{1}{2}}\sigma} e^{-\frac{(x_i-\mu)}{2\sigma}}$$

and the corresponding likelihood

$$\mathrm{L}(\mu,\sigma) = \frac{1}{(2\pi)^{\frac{n}{2}}\sigma^n} e^{-\sum_{i=1}^{n}\frac{(x_i-\mu)}{2\sigma}}$$

and log-likelihood function are

$$\mathrm{l}(\mu,\sigma) = -\frac{1}{2}\ln\left(\frac{1}{2\pi}\right) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2$$

Where

$$\mu = \frac{1}{n}\sum_{i=1}^{n} x_i$$

and

$$\sigma^2 = \frac{1}{n}\sum(x_i-\mu)^2$$

The variance (σ^2) can be decomposed into the variance due to sampling (σ^2) and additional variance (λ^2) due to process error.

Multi-variate normal distribution

If $\vec{x}_1, \vec{x}_2 ..., \vec{x}_n$ is an independent sample from a p-variate normal distribution with mean $\vec{\mu}$ and covariance matrix Σ, then the joint distribution of $\vec{x}_1, \vec{x}_2 ..., \vec{x}_n$ is

$$f(\vec{x}_1, ..., \vec{x}_n \mid \vec{\mu}, \Sigma) = \prod_{i=1}^{n} \frac{1}{(2\pi)^{\frac{p}{2}} |\Sigma|^{\frac{1}{2}}} e^{-\frac{1}{2}\left(\vec{x}_i - \vec{\mu}\right)' \Sigma^{-1}\left(\vec{x}_i - \vec{\mu}\right)}$$

Hence the likelihood

$$\mathrm{L}(\vec{\mu}, \Sigma) = \prod_{\mathrm{i=1}}^{\mathrm{n}} \frac{1}{(2\pi)^{\frac{p}{2}} |\Sigma|^{\frac{1}{2}}} e^{-\frac{1}{2}\sum_{i=1}^{n}\left(\vec{x}_i - \vec{\mu}\right)' \Sigma^{-1}\left(\vec{x}_i - \vec{\mu}\right)}$$

and the log-likelihood functions are

$$\begin{aligned} \mathrm{l}(\vec{\mu}, \Sigma) &= \ln \mathrm{L}(\vec{\mu}, \Sigma) \\ &= -\frac{np}{2} \ln 2\pi - \frac{n}{2} \ln |\Sigma| - \frac{1}{2} \sum_{i=1}^{n} (\vec{x}_i - \vec{\mu})' \Sigma^{-1} (\vec{x}_i - \vec{\mu}) \end{aligned}$$

Where

$$\vec{\mu} = \frac{1}{n} \sum_{i=1}^{n} \vec{x}_i$$

and

$$\hat{\Sigma} = -\frac{1}{n} \left(\vec{x}_i - \vec{\mu}\right)\left(\vec{x}_i - \vec{\mu}\right)'$$

The covariance (Σ) can be decomposed into the variance due to sampling (Ω) and additional variance (Λ) due to process error.

$$\Sigma = \Omega + \Lambda$$

lognormal distribution

In this case, the pdf is

$$f(x_1, ..., x_n \mid \mu, \sigma) = \frac{1}{x\ \sigma\sqrt{2\pi}} e^{\left(\frac{-(\ln(x) - \ln(\mu))^2}{2\sigma^2}\right)}$$

and log-likelihood is

$$l(\theta) = -\frac{n}{2}\left[\ln(2\pi) + 2\ln(\sigma) + 1\right] + \sum_{i=1}^{n} \frac{1}{x_i},$$

Where

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}\left(\ln(x_i) - \mu\right)^2.$$

Multivariate lognormal distribution

If $\vec{x}_1, \vec{x}_2 ..., \vec{x}_n$ is an independent sample from a p-variate lognormal distribution with mean $\vec{\mu}$ and covariance matrix Σ, then the joint distribution of $\vec{x}_1, \vec{x}_2 ..., \vec{x}_n$ is

$$f(\vec{x}_1, ..., \vec{x}_n \mid \vec{\mu}, \Sigma) = \prod_{i=1}^{n} \frac{1}{x_i (2\pi)^{\frac{p}{2}} |\Sigma|^{\frac{1}{2}}} e^{-\frac{1}{2}\left(\log(\vec{x}_i) - \vec{\mu}\right)' \Sigma^{-1}\left(\log(\vec{x}_i) - \vec{\mu}\right)}$$

and

$$E(x_i) = e^{\mu_i + \frac{\sigma_{ii}}{2}}$$

$$Var(x_i) = e^{2\mu_i} e^{\sigma_{ii}}[e^{\sigma_{ii}} - 1].$$

$$Cor\left(x_i, x_j\right) = \exp\left(\mu_i + \mu_j + \frac{1}{2}\left(\sigma_{ii} + \sigma_{jj}\right)\right) \cdot \left(\exp\left(\rho_{ij}\sqrt{\sigma_{ii}\sigma_{jj}}\right) - 1\right)$$

where ρ_{ij} is the correlation between $\log(x_i)$ and $\log(x_j)$.

For more details concerning properties of the multivariate lognormal distribution see Johnson and Kotz (1972).

Hence the likelihood

$$L(\vec{\mu}, \Sigma) = \frac{1}{\prod_{i=1}^{n} x_i} \frac{1}{(2\pi)^{\frac{np}{2}} |\Sigma|^{\frac{np}{2}}} e^{-\frac{1}{2}\sum_{i=1}^{n}\left(\log(\vec{x}_i) - \vec{\mu}\right)' \Sigma^{-1}\left(\log\left(\vec{x}_i\right) - \vec{\mu}\right)}$$

and the log-likelihood functions are

$$l(\vec{\mu}, \Sigma) = -\prod_{i=1}^{n} x_i \frac{np}{2}\ln 2\pi - \frac{n}{2}\ln|\Sigma| - \frac{1}{2}\sum_{i=1}^{n}\left(\log\left(\vec{x}_i\right) - \vec{\mu}\right)' \Sigma^{-1}\left(\log\left(\vec{x}_i\right) - \vec{\mu}\right)$$

Where

$$\vec{\mu} = \frac{1}{n}\sum_{i=1}^{n}\log\left(\vec{x}_i\right)$$

and

$$\hat{\Sigma} = -\frac{1}{n}\left(\log(\vec{x}_i) - \vec{\mu}\right)\left(\log(\vec{x}_i) - \vec{\mu}\right)'$$

The covariance (Σ) can be decomposed into the variance due to sampling (Ω) and additional variance (Λ) due to process error.

$$\Sigma = \Omega + \Lambda$$

Application to stock assessment

There are two main sources of information for most stock assessment methods, namely information on trends in population size (e.g. survey indices and standardised CPUE series) and information on the age structure of catches (Punt and Hilborn 1997). Likelihood formulations for each of these are shown below, presented as full (negative) log-likelihood functions (including all constants) to allow for the use of AIC and similar statistics for model comparisons (Porch 2003, Burnham and Anderson 2002).

<u>Abundance indices</u>

Lognormal likelihood

Assuming that the abundance indices are log-normally distributed about their corresponding model estimates, the negative log-likelihood contribution ($-\ln L_1$) from all series i for which abundance indices are available is as follows:

$$-\ln L_1 = \tfrac{1}{2}\sum_i \mathrm{W}_{1,i} \sum_{y \in Y_{1,i}} \left\{ \ln[2\pi\,(\sigma_{i,y}^2 + \lambda_i^2) X_{i,y}^2] + \frac{[\ln X_{i,y} - \ln(q_i \hat{X}_{i,y})]^2}{\sigma_{i,y}^2 + \lambda_i^2} \right\} \qquad [1]$$

where

$X_{i,y}$	is the abundance index observation for year y and series i;
$\hat{X}_{i,y}$	is the corresponding model estimate for $X_{i,y}$;
q_i	is the constant of proportionality associated with series i;
$\sigma_{i,y}^2 + \lambda_i^2$	is the residual variance for year y and series i, where $\sigma_{i,y}^2$ represents the sampling component of this variance associated with each observation (i.e. each year y) of series i, and λ_i^2 represents the extent of additional variance (over and above that linked to sampling – Punt and Butterworth 2003, Porch 2003) associated with series i;
$Y_{1,i}$	represents the years for which abundance observations are available for series i; and
$\mathrm{W}_{1,i}$	is the relative weight given to the abundance index component of the likelihood associated with series i.

The assumption of log-normality allows the sampling component of the residual standard deviation, $\sigma_{i,y}$, to be approximated by the CV (coefficient of variation) of the untransformed distributions of the $X_{i,y}$.

Estimation options for λ_i and q_i, given $\sigma_{i,y}$ (a fixed input value) are as follows:

(a) λ_i fixed

(i) If $\lambda_i = 0$ then it is necessary that $\sigma_{i,y} > 0$ for all y. Setting $\sigma_{i,y} = 1$ for all y reduces equation [1] to a least-squares formulation.

(ii) If $\lambda_i > 0$ then $\sigma_{i,y} \geq 0$ for all y. Setting $\sigma_{i,y} = \sigma_i$ for all y ($\sigma_i \geq 0$) allows for equal weighting of all observations relative to one another within an abundance series.

If q_i is to be estimated, then a closed form solution exists, as follows:

$$\ln q_i = \frac{\sum_{y \in Y_{1,i}} [\ln X_{i,y} - \ln \hat{X}_{i,y}] / [\sigma_{i,y}^2 + \lambda_i^2]}{\sum_{y \in Y_{1,i}} 1/[\sigma_{i,y}^2 + \lambda_i^2]} \qquad [2]$$

If $\sigma_{i,y} = \sigma_i$ for all y, then equation [2] reduces to the following:

$$\ln q_i = \sum_{y \in Y_{1,i}} [\ln X_{i,y} - \ln \hat{X}_{i,y}] / \sum_{y \in Y_{1,i}} 1 \qquad [3]$$

b) λ_i estimated

(i) $\sigma_{i,y} = 0$ for all y:
As for (a)(ii) above, this allows equal weighting for all observations within an abundance series. A closed-form solution for λ_i exists, as follows:

$$\lambda_i^2 = \sum_{y \in Y_{1,i}} [\ln X_{i,y} - \ln(q_i \hat{X}_{i,y})]^2 / \sum_{y \in Y_{1,i}} 1 \qquad [4]$$

If q_i is also estimated, then the closed-form solution given in equation [3] applies.

(ii) $\sigma_{i,y} > 0$:
In this case, a closed form solution for λ_i is no longer straight forward. An estimate for λ_i can be obtained by treating it as an estimable parameter in the non-linear optimisation of equation [1], or by finding the value for λ_i that satisfies the following equation:

$$f_i = \sum_{y \in Y_{1,i}} \frac{\sigma_{i,y}^2 + \lambda_i^2 + [\ln X_{i,y} - \ln(q_i \hat{X}_{i,y})]^2}{[\sigma_{i,y}^2 + \lambda_i^2]^2} = 0 \qquad [5]$$

If q_i is also to be estimated, then equation [2] can be used in conjunction with equation [5] to simultaneously solve for q_i and λ_i using an iterative algorithm. Alternatively, both q_i and λ_i could be treated as estimable parameters.

<u>Age composition data</u>

Ernst (2002) identified three different approaches for treating age composition data in a likelihood function. All three approaches result in multinomial-type likelihoods. The first (the option usually used) assumes the age composition data have a multinomial distribution about their expected values (Methot 1989, Punt and Hilborn 1997). The second uses a robustified normal likelihood formulation (Fournier et al. 1990, Hilborn et al. 2003). These first two approaches require the specification of an effective sample size (i.e. number of independent sample units). This can prove difficult if age composition data are not based on simple random samples from the total catch, which is often the case (Punt and Kennedy 1997, McAllister and Ianelli 1997).

The third approach avoids arbitrariness in the specification of the effective sample size by assuming a lognormal distribution for the age composition data, where the CV is taken to be inversely proportional to the square-root of the expected value (Punt and Kennedy 1997, Smith and Punt 1997, Ernst 2002). This form has its basis in the mean-variance relationship for multinomial sampling, and allows larger proportions to be given greater weight, so that undue importance is not given to observations based on only a few samples (Punt and Kennedy 1997, Geromont and Butterworth 1999, De Oliveira 2003). Punt (pers. commn) has more recently recommended that for the lognormal formulation, the CV should instead be inversely proportional to the square-root of the <u>observed</u> value. This recommendation is based on the simulation work by Ernst (2002) that showed better performance (in terms of estimation bias) of the robustified normal likelihood when variance was based on observed rather than expected values.

All three likelihood options are considered for proportion-at-age data.

Multinomial Likelihood
Assuming multinomial distribution, the negative log-likelihood contribution (-lnL_2) from all series i for which proportion-at-age data are available is as follows:

$$-\ln L_2 = \sum_i \mathrm{W}_{2,i} \sum_{y \in Y_{2,i}} \left\{ -\ln(n_{i,y}!) + \sum_{a \in A_i} \ln[(n_{i,y} p_{i,y,a})!] - \sum_{a \in A_i} n_{i,y} p_{i,y,a} \ln(\hat{p}_{i,y,a}) \right\} \qquad [6]$$

where

$p_{i,y,a}$	is the proportion-at-age observation for age a, year y and series i;
$\hat{p}_{i,y,a}$	is the corresponding model estimate for $p_{i,y,a}$;
$n_{i,y}$	is the effective sample size associated with series i in year y;
A_i	represents the ages for which proportion-at-age data are available for series i;
$Y_{2,i}$	represents the years for which proportion-at-age data are available for series i; and
$\mathrm{W}_{2,i}$	is the relative weight given to the proportion-at-age component of the likelihood associated with series i.

Robustified Normal Likelihood
Assuming the robustified version of the normal distribution (Fournier *et al.* 1990), the negative log-likelihood contribution (-lnL_2) from all series i for which proportion-at-age data are available is as follows:

$$-\ln L_2 = \sum_i \mathrm{W}_{2,i} \sum_{y \in Y_{2,i}} \left\{ \begin{array}{l} \frac{1}{2} \sum_{a \in A_i} \ln[2\pi(\xi_{i,y,a} + 0.1/n_{A_i})] + \frac{1}{2} n_{A_i} \ln \tau_{i,y}^2 \\ - \sum_{a \in A_i} \ln\left[\exp\left(\frac{-(p_{i,y,a} - \hat{p}_{i,y,a})^2}{2(\xi_{i,y,a} + 0.1/n_{A_i})\tau_{i,y}^2} \right) + 0.01 \right] \end{array} \right\} \qquad [7]$$

where definitions are as for equation [6], and

n_{A_i} is the number of ages in the set A_i;

$\xi_{i,y,a}$ $= p_{i,y,a}(1\text{-}p_{i,y,a})$; and

$\tau^2_{i,y}$ $= 1/n_{i,y}$.

Lognormal Likelihood with modified variance

Assuming a lognormal likelihood with a variance modified to allow proportions based on bigger sample sizes to be given greater weight (Punt and Kennedy 1997), the age data are available is as follows:

$$-\ln L_2 = \tfrac{1}{2}\sum_i \mathrm{W}_{2,i} \sum_{y\in Y_{2,i}} \sum_{a\in A_i} \left\{ \ln[2\pi(\sigma^2_{p_i}/p_{i,y,a})p^2_{i,y,a}] + \frac{(\ln p_{i,y,a} - \ln \hat{p}_{i,y,a})^2}{(\sigma^2_{p_i}/p_{i,y,a})} \right\} \qquad [8]$$

where definitions are as for equation [6], and a closed form solution for $\sigma^2_{p_i}$ exists as follows:

$$\sigma^2_{p_i} = \frac{1}{(\sum_{y\in Y_{2,i}} 1)(\sum_{a\in A_i} 1)} \sum_{y\in Y_{2,i}} \sum_{a\in A_i} p_{i,y,a}(\ln p_{i,y,a} - \ln \hat{p}_{i,y,a})^2 \qquad [9]$$

References

Burnham, K.P. and D.R. Anderson 2002 – Model selection and multimodel inference: a practical information-theoretic approach. Second Edition. Springer-Verlag, New York: [xxvi] + 488pp.

De Oliveira, J.A.A. 2003 – The development and implementation of a joint management procedure for the South African pilchard and anchovy resources. Ph.D. thesis, University of Cape Town, Cape Town: [iv] + 319pp.

Ernst, B. 2002 – An investigation on length-based models used in quantitative population modeling. Ph.D. thesis, University of Washington, Seattle: 150pp.

Fournier, D.A., Sibert, J.R., Majkowski, J. and J. Hampton 1990 – MULTIFAN a likelihood-based method for estimating growth parameters and age composition from multiple length frequency data sets illustrated using data for southern bluefin tuna (Thunnus maccoyii). Can. J. Fish. Aquat. Sci. 47: 301-317.

Geromont, H.F. and D.S. Butterworth 1999 – Operating models proposed for south coast *Merluccius capensis* management procedure robustness trials. Unpublished Report, Marine and Coastal Management, South Africa WG/11/99/D:H:46: 33pp.

Hilborn, R., Maunder, M.[N.], Parma, A.[M.], Ernst, B., Payne, J.[] and P.[J.] Starr 2003 – Coleraine. A generalized age-structured stock assessment model. User's Manual Version 2.0 (Revised May 2003): 58pp.

Johnson, N.L. and Kotz,S. (1972). Distributions in Statistics: Continuous Multivariate Distribution. Wiley, New York.

Methot, R.D. 1989 – Synthetic estimates of historical abundance and mortality for northern anchovy. Am. Fish. Soc. Symp. 6: 66-82.

McAllister, M.K. and J.N. Ianelli 1997 – Bayesian stock assessment using catch-age data and the sampling – importance resampling algorithm. Can. J. Fish. Aquat. Sci. 54: 284-300.

Porch, C.E. 2003 – VPA-2Box. Version 3.01. User's Guide.

Punt, A.E. and D.S. Butterworth 2003 – Specifications and clarifications regarding the ADAPT VPA assessment/projection computations carried out during the September 2000 ICCAT West Atlantic bluefin tuna stock assessment session. Col. Vol. Sci. Pap. ICCAT, 55(3): 1028-1040.

Punt, A.E. and R. Hilborn 1997 – Fisheries stock assessment and decision analysis: the Bayesian approach. Reviews in Fish Biology and Fisheries 7: 35-63.

Punt, A.E. and R.B. Kennedy 1997 – Population modelling of Tasmanian rock lobster, *Jasus edwardsii*, resources. Mar. Freshwater Res. 48: 967-980.

Smith, A.D.M and A.E. Punt 1998 – Stock assessment of gemfish (*Rexea solandri*) in eastern Australia using maximum likelihood and Bayesian methods. In Fishery Stock Assessment Models. F. Funk, T.J. Quinn II, J. Heifetz, J.N. Ianelli, J.E. Powers, J.F. Schweigert, P.J. Sullivan, and C.I. Zhang (Eds). Alaska Sea Grant College Program Report No. AK-SG-98-01, University of Alaska Fairbanks: 245-286.