It's Only Fitting

“Fitting model to data” ≡ “parameterizing model” ≡ estimating unknown parameters in the model

Likelihood: an example

Cohort of $n_0$ individuals; observe survivors at times $t = 1, 2, 3, ..., q$:

$n_1, n_2, ..., n_q$

Suppose we want to “fit” the demographic survival model. The probability of this particular outcome under the model is the joint probability mass function:

$$\Pr[N_1 = n_1, N_2 = n_2, \ldots, N_q = n_q \mid N_0 = n_0] = p(n_1 \mid n_0) p(n_2 \mid n_1) \cdots p(n_q \mid n_{q-1})$$
\[
= \prod_{t=1}^{q} \left( \frac{n_{t-1}}{n_t} \right) p^{n_t} (1 - p)^{n_{t-1} - n_t}
\]

For instance, suppose the numbers of survivors, starting at \( t = 0 \), are: 100, 93, 80, 69, 61, 56, 52, 45, 43, 40, 36. Then

\[
\Pr[N_1 = 93, N_2 = 80, \ldots, N_q = 36 \mid N_0 = 100] = \]

\[
\left( \frac{100}{93} \right) p^{93} (1 - p)^{100-93} \left( \frac{93}{80} \right) p^{80} (1 - p)^{93-80} \ldots \left( \frac{40}{36} \right) p^{36} (1 - p)^{40-36}
\]

\[
= L(p)
\]

R. A. Fisher (1922): use as as estimate of \( p \) the value that maximizes \( L(p) \).
\( L(p) \): likelihood function (joint pmf or pdf, evaluated at the data)

\( \hat{p} \): maximum likelihood (ML) estimate of \( p \)

Likelihoods are usually products, so in maximizing it is usually easier to work with log-likelihoods:

\[
\ln L(p) = \sum_{t=1}^{q} \ln \left( \frac{n_{t-1}}{n_t} \right) + \sum_{t=1}^{q} n_t \ln p + \sum_{t=1}^{q} (n_{t-1} - n_t) \ln p
\]

Set \( d\ln L(p)/dp = 0 \) and solve for \( p \):

\[
\hat{p} = \frac{\sum_{t=1}^{q} n_t}{\sum_{t=1}^{q} n_{t-1}}
\]

For the example data, \( \hat{p} = 0.8998 \)
Hypothesis tests and confidence intervals

Suppose $p_0$ is the true (but unknown) value of $p$

log-likelihood ratio quantity:

$$G^2 = -2[\ln L(p_0) - \ln L(\hat{p})]$$

Math-stat theorem (Wilks 1938):

$G^2$ converges in distribution to a chi-square distribution with 1 df (same type of convergence as CLT)
Result allows for a statistical hypothesis test:

\[ H_0: \ p = p_0 \]
\[ H_1: \ p \neq p_0 \]

Calculate \( G^2 \), reject \( H_0 \) in favor of \( H_1 \) if \( G^2 \) exceeds the \( 100(1 - \alpha) \)th percentile, \( \chi^2_{\alpha}(1) \), of a chi-square (1 df) distribution, for an approximate size \( \alpha \) test.

Result also allows an approximate 100(1 - \( \alpha \))% confidence interval for \( p \) (set of all values of \( p_0 \) for which \( H_0 \) would not be rejected):

\[ \{ p_0: \ln L(\hat{p}) - \ln L(p_0) \leq \chi^2_{\alpha}(1)/2 \} \]

The 95th percentile of a chi-square(1) distribution is about \( \chi^2_{0.05}(1) = 3.843 \). Finding the endpoints of the confidence interval usually requires numerical root-finding. For the data, the 95% CI for \( p \) is (0.8750, 0.9215)
Likelihoods for other models

NLAR model:

\[ X_t = h(X_{t-1}) + E_t \]

Recall that the one-step conditional pdf is

\[ p(x_t \mid x_{t-1}) = \frac{1}{(\sigma^2 2\pi)^{-\frac{1}{2}}} \exp \left[ - \frac{(x_t - h(x_{t-1}))^2}{2\sigma^2} \right] \]

Suppose \( x_0, x_1, \ldots, x_q \) are the (transformed) observations of population abundances.

\( \theta \): vector of unknown parameters in \( h(x_{t-1}) \)

\( \sigma^2 \): variance of noise, \( E_t \)

Likelihood function:

\[ L(\theta, \sigma^2) = \prod_{t=1}^{q} p(x_t \mid x_{t-1}) \]
Log-likelihood:

$$\ln L(\theta, \sigma^2) = \sum_{t=1}^{q} \ln p(x_t | x_{t-1})$$

$$= - (q/2) \ln (2\pi) - (q/2) \ln (\sigma^2) - \left[ 1/(2\sigma^2) \right] \sum_{t=1}^{q} [x_t - h(x_{t-1})]^2$$

ML estimates $\widehat{\theta}, \widehat{\sigma}^2$ are the values of $\theta$ and $\sigma^2$ that maximize $L$ or $\ln L$.

ex. environmental stochastic Ricker model ($X_t = \ln N_t$):

$$X_t = X_{t-1} + a - be^{X_{t-1}} + E_t$$

Data: $n_0, n_1, ..., n_q$
Transformed data: $x_0 = \ln n_0, x_1 = \ln n_1,$ $..., x_q = \ln n_q$

$\theta = [a, b]'$
\[ \ln L(a, b, \sigma^2) = \]
\[ - \left( \frac{q}{2} \right) \ln (2\pi) - \left( \frac{q}{2} \right) \ln (\sigma^2) \]
\[ - \left[ \frac{1}{2\sigma^2} \right] \sum_{t=1}^{q} (x_t - x_{t-1} - a + be^{x_{t-1}})^2 \]

**Multivariate NLAR model (such as LPA):**

\[ X_t = h(X_{t-1}) + E_t \]

\[ p(x_t \mid x_{t-1}) = |\Sigma|^{-\frac{1}{2}} (2\pi)^{-\frac{k}{2}} \times \exp \left[ - (x_t - h_{t-1})' \Sigma^{-1} (x_t - h_{t-1}) / 2 \right] \]

\( k: \) number of state variables in \( X_t \)
\( h_{t-1} = h(x_{t-1}) \)

Likelihood is a product of multivariate normal pdf's; log-likelihood is:

\[ \ln L(\theta, \Sigma) = - q(k/2) \ln (2\pi) - (q/2) \ln |\Sigma| \]
\[ - (1/2) \sum_{t=1}^{q} (x_t - h_{t-1})' \Sigma^{-1} (x_t - h_{t-1}) \]
Properties of ML estimates:

asymptotically unbiased \( E(\hat{\theta}) \rightarrow \theta \) as \( q \rightarrow \infty \)

consistent \( \Pr \left( |\hat{\theta} - \theta| < \epsilon \right) \rightarrow 1 \) as \( q \rightarrow \infty \)

asymptotically efficient

\( V(\hat{\theta}) \rightarrow \) theoretical minimum as \( q \rightarrow \infty \)

asymptotically normal: distribution of \( \hat{\theta} \)

converges to a normal distribution as \( q \rightarrow \infty \)
Conditional least squares estimation

Conditional expected value:

\[ E(X_t | X_{t-1} = x_{t-1}) = h(x_{t-1}; \theta) \]

Conditional sum of squares:

\[ Q(\theta) = \sum_{t=1}^{q} [x_t - h(x_{t-1}; \theta)]^2 \]

Conditional least squares (CLS) estimates of the parameters in \( \theta \) jointly minimize \( Q(\theta) \)
Properties of CLS estimates:

asymptotically unbiased

consistent

asymptotically

robust to mis-specification of distribution of $E_t$

(not efficient)
Estimating $\Sigma$ (ML or CLS)

Fitted model yields series of residuals (or residual vectors):

$$e_t = x_t - \hat{h}(x_{t-1})$$

for $t = 1, 2, ..., q$, where $\hat{h}(x_{t-1})$ is the “skeleton” evaluated at the ML or CLS estimate of $\theta$

Let $D = [e_1, e_2, ..., e_q]$ $(k \times q)$; then

$$\hat{\Sigma} = DD'/q$$
Confidence intervals, hypothesis tests

Profile likelihood (ML estimates)

\[ \theta = \begin{bmatrix} \omega \\ \nu \end{bmatrix} \]

\( \omega \) is a parameter vector of interest; \( \nu \) is a vector of "nuisance" parameters

\( L(\omega, \nu, \Sigma) \): likelihood function

\( L\left(\omega, \hat{\nu}_\omega, \hat{\Sigma}_\omega\right) \): likelihood function

maximized for fixed \( \omega \) (profile likelihood function of \( \omega \))

\( L\left(\hat{\omega}, \hat{\nu}, \hat{\Sigma}\right) \): maximized likelihood function

(unrestricted)
Approx 100(1 − α)% confidence region for ω is the set of all values of ω for which

\[
G^2 = -2 \left[ \ln L(\omega, \hat{\nu}_\omega, \hat{\Sigma}_\omega) - \ln L(\hat{\omega}, \hat{\nu}, \hat{\Sigma}) \right] \leq \chi^2_\alpha(\eta)
\]

where η is the number of parameters in ω

**Bootstrapping (ML or CLS estimates)**

\(e_1, e_2, ..., e_q\): residuals (or residual vectors)

\(e_1^*, e_2^*, ..., e_q^*\): random sample with replacement from the centered residuals (i.e. subtract the average from each of \(e_1, e_2, ..., e_q\) to center them around zero)
Generate a “bootstrapped data set” \( x_0, x_1^*, x_2^*, ..., x_q^* \) (simulated data set from estimated model) using

\[
x_t^* = \hat{h}(x_{t-1}^*) + e_t^*
\]

Fit the model to \( x_0, x_1^*, x_2^*, ..., x_q^* \), obtaining bootstrapped parameter estimates \( \theta^*, \Sigma^* \)

Repeat 2000 or so times (using residuals & estimates from original fitted model), obtaining 2000 sets of bootstrapped parameter estimates (or test statistics, or functions of parameter estimates such as eigenvalues or Lyaponov exponents)
Resulting 2000 values constitute a good (statistically consistent) estimate of the variability of the estimates; e.g., 2.5th and 97.5th percentiles of the bootstrapped estimates of a parameter are a valid approximate 95% confidence interval for the parameter.
Model evaluation

Goodness of fit

If the model fits well, then the residuals $e_1, e_2, \ldots, e_q$ should behave approximately like independent, identically distributed normal (or multivariate normal) variates

Evaluating model fit revolves around the residuals; e.g. autocorrelation tests, residual plots to look for unchanging variance, normal probability plots, etc.

“R-square”: proportion of variability in data accounted for by fitted model

Validation (predictive ability of the model)

Ideally, collect new data, independently from the data used to fit the model (possibly with experimental, model-suggested perturbations!)