SAS proc calis: The basics

STA431 Winter/Spring 2013

See last slide for copyright information.
Overview

1. The program
2. Maximum likelihood
3. Goodness of fit test
4. What we get
What it is and what it does

- **SAS proc calis** is model fitting software.
- It fits classical structural equation models to data, using numerical maximum likelihood (or optionally, other methods).
- Most of the output is about the details of the numerical search and how well the model fits.
- This is a narrow focus, compared to most other SAS procedures.
- Still, SAS tells you more than you need or want to know — as usual.
Three programs

- `proc calis` incorporates three programs that originated outside of SAS.
- They all use different, unrelated syntax for specifying the model.
- We will use the `lineqs`\(^2\) syntax, which is the most convenient.

- First, read and label the data as usual in a SAS `data step`.

Specifying the model
Using lineqs syntax

Input includes:

- Names of the observable variables.
- Model equations, pretty much as you would write them by hand
  - Including the regression coefficients and the error terms – you name them.
- No intercepts: The model is given in centered form and SAS bases everything on the sample covariance matrix.
- Naming rules: Names of latent variables (including error terms) must begin with the letter F, D or E.
- Names must also be given to the variances and covariances of the explanatory variables and error terms. Anything unspecified is assumed zero.
- In the end, you give names to all the non-zero parameters in your model.
What happened to the intercepts?

$$L(\mu, \Sigma) = |\Sigma|^{-n/2}(2\pi)^{-np/2} \exp -\frac{n}{2} \left\{ \text{tr}(\hat{\Sigma}\Sigma^{-1}) + (\bar{x} - \mu)'\Sigma^{-1}(\bar{x} - \mu) \right\}$$

- Remember, $\mu$ and $\Sigma$ are both functions of $\theta$.
- For regression without measurement error, expected values and intercepts are identifiable, but if there are latent variables that’s rare.
- Re-parameterize, absorbing expected values and intercepts into $\mu$.
- Estimate $\mu$ with $\bar{x}$ and it’s gone.
- This is just a technical trick to allow the likelihood to have a unique maximum.
- But it does no harm, because *relationships* between variables are represented by the covariances.
Maximum likelihood

\[
L(\Sigma) = |\Sigma|^{-n/2}(2\pi)^{-np/2} \exp - \frac{n}{2} \left\{ tr \left( \hat{\Sigma} \Sigma^{-1} \right) \right\}
\]

\[
L_2(\theta) = |\Sigma(\theta)|^{-n/2}(2\pi)^{-np/2} \exp - \frac{n}{2} \left\{ tr \left( \hat{\Sigma} \Sigma(\theta)^{-1} \right) \right\}
\]

- Can maximize \( L(\Sigma) \) over all \( \Sigma \in \mathcal{M} \), or maximize \( L_2(\theta) \) over all \( \theta \in \Theta \).
- If the function connecting \( \Sigma \) and \( \theta \) is one-to-one and there is the same number of \( \theta \) and unique \( \Sigma \) values, call the parameter \( \theta \) just identifiable.
- In this case it’s the same problem, and
- The invariance principle can be used to go back and forth between \( \hat{\Sigma} \) and \( \hat{\theta} \).
- Otherwise . . .
Maximize $L_2(\theta)$ over all $\theta \in \Theta$

$$L_2(\theta) = |\Sigma(\theta)|^{-n/2} (2\pi)^{-np/2} \exp \left[ - \frac{n}{2} \left\{ tr \left( \hat{\Sigma} \Sigma(\theta)^{-1} \right) \right\} \right]$$

- Actually, maximize the log likelihood.
- Well, actually, minimize the minus 2 log likelihood.
- Well, actually, minimize the minus 2 log likelihood plus a carefully chosen constant.

- The constant is based on the likelihood ratio test for goodness of model fit.
Likelihood ratio tests

In general

Setup

\[ Y_1, \ldots, Y_n \stackrel{i.i.d.}{\sim} P_\theta, \ \theta \in \Theta, \]
\[ H_0 : \theta \in \Theta_0 \subset \Theta \ \text{v.s.} \ H_1 : \theta \in \Theta_1 = \Theta \cap \Theta_0^c \]

Test Statistic:

\[ G^2 = -2 \ln \left( \frac{\max_{\theta \in \Theta_0} L(\theta)}{\max_{\theta \in \Theta} L(\theta)} \right) \]
What to do
And how to think about it

\[ G^2 = -2 \ln \left( \frac{\max_{\theta \in \Theta_0} L(\theta)}{\max_{\theta \in \Theta} L(\theta)} \right) \]

- Maximize the likelihood over the whole parameter space. You already did this to calculate the MLE. Evaluate the likelihood there. That’s the denominator.
- Maximize the likelihood over just the parameter values where \( H_0 \) is true – that is, over \( \Theta_0 \). This yields a restricted MLE. Evaluate the likelihood there. That’s the numerator.
- The numerator cannot be larger, because \( \Theta_0 \subset \Theta \).
- If the numerator is a lot less than the denominator, the null hypothesis is unbelievable, and
  - The ratio is close to zero
  - The log of the ratio is a big negative number
  - \(-2\) times the log is a big positive number
  - Reject \( H_0 \) when \( G^2 \) is large enough.
Distribution of $G^2$ when $H_0$ is true

Given some technical conditions,

- $G^2$ has an approximate chi-squared distribution under $H_0$ for large $n$.
- Degrees of freedom equal number of (non-redundant) equalities specified by $H_0$.
- Reject $H_0$ when $G^2$ is larger than the chi-squared critical value.
Goodness of fit test for a covariance structure model
Multivariate normal data

Call it a “covariance structure” model because $\Sigma = \Sigma(\theta)$.

- Compare fit of model to fit of the best possible model.
- The best possible model is the unrestricted multivariate normal:
  - Estimate $\mu$ with $\bar{x}$.
  - Estimate $\Sigma$ with $\hat{\Sigma}$.
- Covariance structure model is re-parameterized to get rid of intercepts, so again, $\mu$ is estimated with $\bar{x}$.
- Compare
  \[
  \ln L \left( \hat{\Sigma} \right) \text{ to } \ln L \left( \Sigma(\hat{\theta}) \right)
  \]
Likelihood ratio test
For goodness of model fit

Difference in fit (times two):

\[ G^2 = 2 \left( \ln L \left( \hat{\Sigma} \right) - \ln L \left( \Sigma(\hat{\theta}) \right) \right) \]

\[ = -2 \ln \left( \frac{L \left( \Sigma(\hat{\theta}) \right)}{L \left( \hat{\Sigma} \right)} \right) \]

It looks like a likelihood ratio test statistic.
More details

\[ G^2 = -2 \ln \left( \frac{L(\Sigma(\hat{\theta}))}{L(\hat{\Sigma})} \right) \]

If the covariance structure model is correct and

- The parameter vector is identifiable, and
- There are more unique variances and covariances in \( \Sigma \) than there are model parameters in \( \theta \), and
- Some other technical conditions hold

Then for large samples, \( G^2 \) has an approximate chi-squared distribution, with degrees of freedom the number of variances-covariances \textit{minus} the number of model parameters.
Simplify $G^2 = 2 \left( \ln L \left( \hat{\Sigma} \right) - \ln L \left( \Sigma(\hat{\theta}) \right) \right)$

Recalling $L(\Sigma) = |\Sigma|^{-n/2} (2\pi)^{-np/2} \exp \left\{ -\frac{n}{2} \left\{ tr \left( \hat{\Sigma} \Sigma^{-1} \right) \right\} \right\}$,

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

$= n \left( tr(\hat{\Sigma} \Sigma(\hat{\theta})^{-1}) + \ln |\Sigma(\hat{\theta})| - \ln |\hat{\Sigma}| - p \right)$
A cute way to maximize the likelihood over $\theta \in \Theta$

- Minimize $G^2(\theta)$: Just $-2$ log likelihood plus a constant.

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

$$= n \left( \text{tr}(\hat{\Sigma}\Sigma(\theta)^{-1}) + \ln |\Sigma(\theta)| - \ln |\hat{\Sigma}| - p \right)$$

- Actually, minimize the “Objective Function”

$$\text{tr}(\hat{\Sigma}\Sigma(\theta)^{-1}) + \ln |\Sigma(\theta)| - \ln |\hat{\Sigma}| - p$$

- Multiply by $n$ (or $n - 1$) to get the $G^2$ statistic.

- This is what SAS proc calis does.
### Saturated models
All the degrees of freedom in the data are “soaked up” by the model.

- If there are the same number of moment structure equations and unknown parameters and the parameter is identifiable, there is a one-to-one function between $\hat{\Sigma}$ and $\hat{\theta}$.
- In this case the parameter is called *just identifiable*.
- $L\left(\hat{\Sigma}\right) = L\left(\Sigma(\hat{\theta})\right)$
- $G^2 = 0, df = 0$ and the standard test for goodness of fit does not apply.
- The model may still be testable some other way.
What does `proc calis` give us?

- An indication of whether the numerical search went okay.
- MLEs of all the parameters, standard errors and $Z$ tests of $H_0 : \theta_j = 0$.
- The $-2$ log likelihood at the MLE, plus a constant.
- Likelihood ratio test for goodness of fit.
<table>
<thead>
<tr>
<th>The program</th>
<th>Maximum likelihood</th>
<th>Goodness of fit test</th>
<th>What we get</th>
</tr>
</thead>
</table>

With the -2 log likelihood at the MLE (plus a constant) we can

- Fit a full and a reduced model.
- Test null hypothesis that the reduced model holds, using a LR test.
- $G^2$ is a difference between two -2 log likelihoods.
- The constant ($-2 \ln L(\hat{\Sigma})$) cancels.

- This is all we really need.
Copyright Information

This slide show was prepared by Jerry Brunner, Department of Statistical Sciences, University of Toronto. It is licensed under a Creative Commons Attribution - ShareAlike 3.0 Unported License. Use any part of it as you like and share the result freely. The \LaTeX source code is available from the course website:
http://www.utstat.toronto.edu/~brunner/oldclass/431s31