Chapter 6 - Model selection and checking

02433 - Hidden Markov Models

Rune H. B. Christensen, Martin Wæver Pedersen, Henrik Madsen

Course week 6

MWP, compiled June 7, 2011

Lunds universitet
Lunds Tekniska Högskola

Interreg IVA
ÖRESUND – KATTEGAT – SKAGERRAK

DTU

Model selection

Purpose of model selection: Identify the "best" model (in some sense. . .).

Two commonly applied criteria:

- Akaike's information criteria: $\mathrm{AIC} = -2\log L + 2p$.
- Bayesian information criteria: $\mathrm{BIC} = -2\log L + p\log N$,

where L is the likelihood of the fitted model, p is the number of parameters, and N is the number of data points. The term $-2\log L$ decreases with increasing model complexity (more parameters), whereas the penalties $2p$ or $p\log N$ increase with increasing complexity. The BIC applies a larger penalty when $N > e^2 = 7.4$.

Model selection: The lower the AIC/BIC value the better the model (only compare AIC with AIC and BIC with BIC values!).

Earthquake example

Here is listed the AIC and BIC of four different models fitted to the earthquake data:

Model	model-df	$-\log L$	AIC	BIC
indep. Poisson	1	391.91	785.8	788.5
2-state HMM	4	342.32	692.6	703.3
3-state HMM	9	329.46	**676.9**	**701.0**
4-state HMM	16	327.83	687.7	730.4

The 3-state HMM has the lowest criteria values, and therefore seems to be the best model with respect to both AIC and BIC. A Poisson model assuming independent observations is clearly not appropriate. The 4-state HMM, while having a lower negative log-likelihood is punished for its complexity.

Pseudo-residuals for continuous observations (1/2)

In many branches of statistical modelling (e.g. regression, glm) it is common to use residuals as a means to check the validity of the fitted model. In the context of these models the residuals are calculated from the expected observation and the actual observation. Pseudo-residuals as presented here are a similar concept, however with general applicability to HMMs.

Uniform pseudo-residual for the observation x_t is

$$u_t = \Pr(X_t \leq x_t) = F_{X_t}(x_t),$$

where X_t is a continuous random variable and F is its cumulative distribution function (c.d.f.).

Normal pseudo-residual for the observation x_t is

$$z_t = \Phi^{-1}(u_t) = \Phi^{-1}(F_{X_t}(x_t)),$$

where Φ is the c.d.f. of a standard normal distributed random variable.

Pseudo-residuals for continuous observations (2/2)

Uniform pseudo-residuals
The uniform pseudo-residuals are uniformly distributed if the model is correct. They can therefore be visualized in a histogram or qq-plot to check the validity of the model. The uniform pseudo-residuals are, however, not well suited for outlier identification since the distribution has a bounded support on $[0, 1]$.

Normal pseudo-residuals
The normal pseudo-residuals are standard normal distributed if the model is correct. They have similar properties to uniform pseudo-residuals with respect to model validation. Normal pseudo-residuals are furthermore well suited for outlier detection since their absolute value indicates the deviation from the median of the distribution.

Pseudo-residual for discrete observations (1/2)

The pseudo-residuals for discrete observations (Poisson, binomial etc.) are specified slightly differently namely as an interval (or segment) rather than a value.

Define x_t^- as the largest possible realization of $X_t | X_t < x_t$. For example for a Poisson distribution having $x_t = 3$, then $x_t^- = 2$.

Uniform pseudo-residual segments are:

$$[u_t^-; u_t^+] = [F_{X_t}(x_t^-); F_{X_t}(x_t)],$$

where x_t maps to u_t^+.

Normal pseudo-residual segments are:

$$[z_t^-; z_t^+] = [\Phi^{-1}(u_t^-); \Phi^{-1}(u_t^+)].$$

Pseudo-residual for discrete observations (2/2)

If one wishes to use the discrete pseudo-residuals for model validation e.g. in a qq-plot, a way to sort the residuals is required. This is done by calculating so-called "mid-pseudo-residuals" as

$$z_t^m = \Phi^{-1}\left(\frac{u_t^- + u_t^+}{2}\right).$$

Types of pseudo-residuals (1/2)

The **ordinary** pseudo-residual for x_t is based on its conditional distribution given all other data, i.e. for continuous variables:

$$z_t = \Phi^{-1}\left(P(X_t \leq x_t | \mathbf{X}^{(-t)} = \mathbf{x}^{(-t)})\right),$$

and for discrete variables

$$z_t^- = \Phi^{-1}\left(P(X_t < x_t | \mathbf{X}^{(-t)} = \mathbf{x}^{(-t)})\right),$$

$$z_t^+ = \Phi^{-1}\left(P(X_t \leq x_t | \mathbf{X}^{(-t)} = \mathbf{x}^{(-t)})\right).$$

The **forecast** pseudo-residual for x_t is based on its conditional distribution given data prior to t, i.e. for continuous variables

$$z_t = \Phi^{-1}\left(P(X_t \leq x_t | \mathbf{X}^{(t-1)} = \mathbf{x}^{(t-1)})\right),$$

and for discrete variables

$$z_t^- = \Phi^{-1}\left(P(X_t < x_t | \mathbf{X}^{(t-1)} = \mathbf{x}^{(t-1)})\right),$$

$$z_t^+ = \Phi^{-1}\left(P(X_t \leq x_t | \mathbf{X}^{(t-1)} = \mathbf{x}^{(t-1)})\right).$$

Types of pseudo-residuals (2/2)

The **ordinary** pseudo-residuals
The general means to detect deviating observations based on the information in the remaining dataset. These are commonly used to check the model if it is to be used for inference within the data.

The **forecast** pseudo-residuals
These are called forecast pseudo-residuals since they rely on the one-step forecast distribution of x_t, i.e. they are computed based on the data observed prior to t. These are commonly used to evaluate the model's forecasting ability.

Exercises

1,4,5