Analysis of Time Series, L17

Rolf Larsson

Uppsala University

14 maj 2025

Today

Preparation:

- Appendix B: Multivariate normal distribution
- The companion form

Chap. 6:

- State-Space Models
- The Kalman Filter
- The EM algorithm
- Maximum Likelihood

Multivariate normal distribution

p.497: Let $\mathbf{y} = (y_1, ..., y_m)'$, $\mathbf{x} = (x_1, ..., x_n)'$ and suppose that

$$\begin{pmatrix} \mathbf{y} \\ \mathbf{x} \end{pmatrix} \sim N \left\{ \begin{pmatrix} \boldsymbol{\mu}_y \\ \boldsymbol{\mu}_x \end{pmatrix}, \begin{pmatrix} \Sigma_{yy} & \Sigma_{yx} \\ \Sigma_{xy} & \Sigma_{xx} \end{pmatrix} \right\}$$

.

Then, $\mathbf{y}|\mathbf{x} \sim N\left(\boldsymbol{\mu}_{y|x}, \Sigma_{y|x}\right)$ with

$$\begin{aligned} \boldsymbol{\mu}_{y|x} &= \boldsymbol{\mu}_y + \Sigma_{yx}\Sigma_{xx}^{-1}(\mathbf{x} - \boldsymbol{\mu}_x), \\ \Sigma_{y|x} &= \Sigma_{yy} - \Sigma_{yx}\Sigma_{xx}^{-1}\Sigma_{xy}. \end{aligned}$$

The companion form

The AR(p) model

$$x_t = \phi_1 x_{t-1} + ... + \phi_p x_{t-p} + w_t$$

may be written in terms of one lag as

$$\begin{pmatrix} x_t \\ x_{t-1} \\ \vdots \\ x_{t-p+1} \end{pmatrix} = \begin{pmatrix} \phi_1 & \phi_2 & \cdots & & \phi_p \\ 1 & 0 & \cdots & & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} x_{t-1} \\ x_{t-2} \\ \vdots \\ x_{t-p} \end{pmatrix} + \begin{pmatrix} w_t \\ 0 \\ \vdots \\ 0 \end{pmatrix},$$

or in short, $\mathbf{x}_t = \mathbf{\Phi}\mathbf{x}_{t-1} + \mathbf{w}_t$.

This is also called *the companion form*.

State-Space Models

- General model

$$\mathbf{y}_t = A_t \mathbf{x}_t + \Gamma \mathbf{u}_t + \mathbf{v}_t, \quad \text{(observation equation)}$$
$$\mathbf{x}_t = \Phi \mathbf{x}_{t-1} + \Upsilon \mathbf{u}_t + \mathbf{w}_t, \quad \text{(state equation)}$$

where $E(\mathbf{v}_t \mathbf{v}_t') = R$, $E(\mathbf{w}_t \mathbf{w}_t') = Q$.

- Example 1: MA(1) with zero mean

$$y_t = \begin{pmatrix} 1 & \theta \end{pmatrix} \begin{pmatrix} w_t \\ w_{t-1} \end{pmatrix},$$
$$\begin{pmatrix} w_t \\ w_{t-1} \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} w_{t-1} \\ w_{t-2} \end{pmatrix} + \begin{pmatrix} w_t \\ 0 \end{pmatrix},$$

i.e. $A_t = \begin{pmatrix} 1 & \theta \end{pmatrix}$, $\mathbf{x}_t = \begin{pmatrix} w_t \\ w_{t-1} \end{pmatrix}$, $\Phi = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}$, $\mathbf{w}_t = \begin{pmatrix} w_t \\ 0 \end{pmatrix}$,

$\Gamma = 0$, $\mathbf{v}_t = 0$, $R = 0$, $\Upsilon = 0$, $Q = \sigma_w^2 \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$.

State-Space Models

General:

- $$\begin{aligned} \mathbf{y}_t &= A_t \mathbf{x}_t + \Gamma \mathbf{u}_t + \mathbf{v}_t, \quad \text{(observation equation)} \\ \mathbf{x}_t &= \Phi \mathbf{x}_{t-1} + \Upsilon \mathbf{u}_t + \mathbf{w}_t. \quad \text{(state equation)} \end{aligned}$$
- The data up to time s is $Y_s = \{\mathbf{y_1}, ..., \mathbf{y_s}\}$.
- Conditional expectation
$$\mathbf{x}_t^s = E(\mathbf{x}_t | Y_s)$$
- Mean square error matrix
$$P_t^s = E\{(\mathbf{x}_t - \mathbf{x}_t^s)(\mathbf{x}_t - \mathbf{x}_t^s)'\}$$
- *Forecasting* when $s < t$, *filtering* when $s = t$, *smoothing* when $s > t$.

The Kalman Filter

Theorem (Property 6.1 ($\Upsilon = \Gamma = 0$))

$$\begin{aligned}
\mathbf{x}_t^{t-1} &= \Phi \mathbf{x}_{t-1}^{t-1}, \\
P_t^{t-1} &= \Phi P_{t-1}^{t-1} \Phi' + Q,
\end{aligned}$$

where

$$\begin{aligned}
\mathbf{y}_t^{t-1} &= A_t \mathbf{x}_t^{t-1}, \\
\Sigma_t &= E\{(\mathbf{y}_t - \mathbf{y}_t^{t-1})(\mathbf{y}_t - \mathbf{y}_t^{t-1})'\} = A_t P_t^{t-1} A_t' + R, \\
K_t &= P_t^{t-1} A_t' (A_t P_t^{t-1} A_t' + R)^{-1}, \\
\mathbf{x}_t^t &= \mathbf{x}_t^{t-1} + K_t(\mathbf{y}_t - A_t \mathbf{x}_t^{t-1}), \\
P_t^t &= (I - K_t A_t) P_t^{t-1}.
\end{aligned}$$

The Kalman Filter

Example 1: MA(1) with zero mean

- $$y_t = \begin{pmatrix} 1 & \theta \end{pmatrix} \begin{pmatrix} w_t \\ w_{t-1} \end{pmatrix},$$
  $$\begin{pmatrix} w_t \\ w_{t-1} \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} w_{t-1} \\ w_{t-2} \end{pmatrix} + \begin{pmatrix} w_t \\ 0 \end{pmatrix},$$
  i.e. $A_t = \begin{pmatrix} 1 & \theta \end{pmatrix}$, $\mathbf{x}_t = \begin{pmatrix} w_t \\ w_{t-1} \end{pmatrix}$, $\Phi = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}$, $\mathbf{w}_t = \begin{pmatrix} w_t \\ 0 \end{pmatrix}$,
  $\mathbf{v}_t = 0$, $R = 0$, $Q = \sigma_w^2 \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$.
- Calculate forecasts $y_1^0, y_2^1, \ldots$ and variances of forecast errors, $\Sigma_1, \Sigma_2, \ldots$.

The EM algorithm

General:

- We observe $\mathbf{y} = (y_1, y_2, ..., y_n)$.
- These are "proxies" for the (partially) unobserved $\mathbf{x} = (x_1, x_2, ..., x_n)$, which is a sample from a distribution with parameter θ.
- Estimate θ.
- Initial estimate $\hat{\theta}_0$.
- The EM algorithm:
  E1: $\hat{\mathbf{y}}_1 = E(\mathbf{x}; \theta = \hat{\theta}_0)$
  M1: $\hat{\theta}_1$ is the MLE based on $\hat{\mathbf{y}}_1$.
  E2: $\hat{\mathbf{y}}_2 = E(\mathbf{x}; \theta = \hat{\theta}_1)$
  M2: $\hat{\theta}_2$ is the MLE based on $\hat{\mathbf{y}}_2$.
  ... Repeat until convergence.
- May be shown to give the MLE of θ under mild regularity conditions.

The EM algorithm

Example 2:

- We observe $y_1 = 1$, $y_2 = 2$, $y_3 = 6$ from a truncated Exponential distribution with parameter (intensity) λ.
  (Corresponds to x_1, x_2, x_3 from non truncated dist.)
- The distribution is truncated at 6.
- $X \sim \mathrm{Exp}(\lambda) \Rightarrow E(X|X > a) = a + \frac{1}{\lambda}$ (why?)
- Estimate λ.
- MLE for non truncated data: $\hat{\lambda} = 1/\bar{y} = 3/\sum_j y_j$.
- Initial estimate $\hat{\lambda}_0 = 3/(1+2+6) = 1/3 \approx 0.3333$.
- *The EM algorithm*:
  E1: $\hat{y}_3 = E(x_3; \lambda = 1/3) = 6 + 1/(1/3) = 9$
  M1: $\hat{\lambda}_1 = 3/(1+2+9) = 1/4 = 0.2500$
  E2: $\hat{y}_3 = E(x_3; \lambda = 1/4) = 6 + 1/(1/4) = 10$
  M2: $\hat{\lambda}_2 = 3/(1+2+10) = 3/13 = 0.2308$
  ... Going on similarly yields $\hat{\lambda}_3 = 9/40 = 0.2250$, $\hat{\lambda}_4 = 27/121 = 0.2231$, $\hat{\lambda}_5 = 81/364 = 0.2225$, $\hat{\lambda}_6 = 243/1093 = 0.2223$, ...

MLE via the EM algorithm

- Suppose that $t = 1, 2, ...n$,
  $\mathbf{y}_t = A_t \mathbf{x}_t + \mathbf{v}_t$, where $\mathbf{v}_t$ is normal with mean 0 and cov. matrix R,
  $\mathbf{x}_t = \Phi \mathbf{x}_{t-1} + \mathbf{w}_t$, where $\mathbf{w}_t$ is normal with mean 0 and cov. matrix Q.
- Parameter vector Θ.
- Likelihood if the $\mathbf{x}_t$ are observed

$$L(\Theta) = f_{\mu_0,\Sigma_0}(\mathbf{x}_0) \prod_{t=1}^{n} f_{\Phi,Q}(\mathbf{x}_t|\mathbf{x}_{t-1}) \prod_{t=1}^{n} f_R(\mathbf{y}_t|\mathbf{x}_t).$$

- From the normal density function,

$$\begin{aligned} -2\log L(\Theta) \propto\ & \log|\Sigma_0| + (\mathbf{x}_0 - \boldsymbol{\mu}_0)'\Sigma_0^{-1}(\mathbf{x}_0 - \boldsymbol{\mu}_0) \\ & + n\log|Q| + \sum_{t=1}^{n}(\mathbf{x}_t - \Phi\mathbf{x}_{t-1})'Q^{-1}(\mathbf{x}_t - \Phi\mathbf{x}_{t-1}) \\ & + n\log|R| + \sum_{t=1}^{n}(\mathbf{y}_t - A_t\mathbf{x}_t)'R^{-1}(\mathbf{y}_t - A_t\mathbf{x}_t). \end{aligned}$$

MLE via the EM algorithm

E step, under assumed parameter values (p.314-316):

$$
\begin{aligned}
E\left\{-2\log L(\Theta)\right\} \propto\ & \log|\Sigma_0| + \mathrm{tr}\left[\Sigma_0^{-1}\left\{P_0^n + (\mathbf{x}_0^n - \boldsymbol{\mu}_0)(\mathbf{x}_0^n - \boldsymbol{\mu}_0)'\right\}\right] \\
& + n\log|Q| + \mathrm{tr}\left[Q^{-1}(S_{11} - S_{10}\Phi' - \Phi S_{10}' + \Phi S_{00}\Phi')\right] \\
& + n\log|R| \\
& + \mathrm{tr}\left[R^{-1}\sum_{t=1}^{n}\left\{(\mathbf{y}_t - A_t\mathbf{x}_t^n)(\mathbf{y}_t - A_t\mathbf{x}_t^n)' + A_tP_t^nA_t'\right\}\right],
\end{aligned}
$$

$$
\begin{aligned}
S_{11} &= \sum_{t=1}^{n}(\mathbf{x}_t^n\mathbf{x}_t^{n\prime} + P_t^n), \\
S_{10} &= \sum_{t=1}^{n}(\mathbf{x}_t^n\mathbf{x}_{t-1}^n{}' + P_{t,t-1}^n), \\
S_{00} &= \sum_{t=1}^{n}(\mathbf{x}_{t-1}^n\mathbf{x}_{t-1}^n{}' + P_{t-1}^n).
\end{aligned}
$$

MLE via the EM algorithm

M step (minimize the expression on the previous slide):

$$\begin{aligned}
\hat{\Phi} &= S_{10}S_{00}^{-1}, \\
\hat{Q} &= n^{-1}(S_{11} - S_{10}S_{00}^{-1}), \\
\hat{R} &= n^{-1}\sum_{t=1}^{n}\left\{(\mathbf{y}_t - A_t\mathbf{x}_t^n)(\mathbf{y}_t - A_t\mathbf{x}_t^n)' + A_tP_t^nA_t'\right\}
\end{aligned}$$

$$\begin{aligned}
\hat{\boldsymbol{\mu}}_0 &= \mathbf{x}_0^n, \\
\hat{\Sigma}_0 &= P_0^n.
\end{aligned}$$

This gives new assumed parameter values in the next E step.
Repeat until convergence.

News of today

- State-Space Models
- The Kalman Filter
  - Forecasting
  - Smoothing
  - MLE via the EM algorithm