

02427 Advanced Time Series Analysis

Lecture 5

Tobias K. S. Ritschel

Department of Applied Mathematics and Computer Science,
Technical University of Denmark

October 1, 2025

Outline

Maximum likelihood estimation

Maximum likelihood estimation (stationary)

Doubly stochastic processes

Extended Kalman filter

Extended Kalman filter for parameter estimation

Generalized state space models

Maximum likelihood estimation

Maximum likelihood estimation

- ▶ Linear stochastic state space model

$$X_t = A_t X_{t-1} + B_t u_{t-1} + e_{1,t}, \quad (1)$$

$$Y_t = C_t X_t + e_{2,t}, \quad (2)$$

$\{e_{1,t}\}$ and $\{e_{2,t}\}$ are mutually uncorrelated normal distributed white noise sequences with variance $\Sigma_{1,t}$ and $\Sigma_{2,t}$

- ▶ Assume that $\dim Y = m$ and that the model is **global identifiable**.
- ▶ The potentially time-varying matrices A_t , B_t , C_t , $\Sigma_{1,t}$ and $\Sigma_{2,t}$ depend on the parameters
- ▶ A missing observation at time t can be handled by

$$\Sigma_{2,t} = \infty \quad (3)$$

ML estimates of state space models

θ are the unknown parameters, $\mathcal{Y}_{\mathcal{N}} = (Y_{\mathcal{N}}, \dots, Y_0)$ are the available observations in the time series of length N ($\mathcal{N} < N$).

The *conditional likelihood function* (conditioned on Y_0) is

$$L(\theta; \mathcal{Y}_{\mathcal{N}}) = f(\mathcal{Y}_{\mathcal{N}} | \theta) \tag{4}$$

$$= f(Y_{\mathcal{N}} | \mathcal{Y}_{\mathcal{N}-1}, \theta) f(Y_{\mathcal{N}-1} | \mathcal{Y}_{\mathcal{N}-2}, \theta) \cdots f(Y_1 | Y_0, \theta). \tag{5}$$

ML estimation for state space models

- ▶ Since the model is linear and all the random variables are Gaussian, $Y_{t+1}|\mathcal{Y}_t$ is Gaussian too
- ▶ This conditional Gaussian density is characterized by the conditional mean and the conditional variance
- ▶ One-step prediction error and its covariance

$$\tilde{Y}_{t+1|t} = Y_{t+1} - \hat{Y}_{t+1|t}, \quad (6)$$

$$\begin{aligned} R_{t+1} &= \text{Var} [\tilde{Y}_{t+1|t}] = \text{Var} [Y_{t+1}|\mathcal{Y}_t] = \Sigma_{t+1|t}^{yy} \\ &= C_t \Sigma_{t+1|t}^{xx} C_t^T + \Sigma_{2,t}. \end{aligned} \quad (7)$$

- ▶ The conditional mean and covariance are simply calculated by an ordinary Kalman Filter

ML estimates

- Conditional density function

$$f(Y_{t+1}|\mathcal{Y}_t) = [(2\pi)^m \det R_{t+1}]^{-\frac{1}{2}} \exp \left[-\frac{1}{2} \tilde{Y}_{t+1|t}^T R_{t+1}^{-1} \tilde{Y}_{t+1|t} \right] \quad (8)$$

- Conditional likelihood function

$$L(\theta; \mathcal{Y}_{\mathcal{N}}) = \prod_{i=1}^{\mathcal{N}} [(2\pi)^m \det R_i]^{-\frac{1}{2}} \exp \left[-\frac{1}{2} \tilde{Y}_{i|i-1}^T R_i^{-1} \tilde{Y}_{i|i-1} \right], \quad (9)$$

- Starting values: $\hat{X}_{0|0} = 0$ and $\Sigma_{0|0}^{xx} = \alpha I$ where α is 'large'. Alternatively, estimate the starting values using the ML approach

ML estimates of state space models

- Alternative: Maximize the log-likelihood

$$\log L(\theta; \mathcal{Y}_{\mathcal{N}}) = -\frac{1}{2} \sum_{i=1}^{\mathcal{N}} [\log \det R_i + \tilde{Y}_{i|i-1}^T R_i^{-1} \tilde{Y}_{i|i-1}] + \text{const.} \quad (10)$$

- Maximum likelihood estimate of θ

$$\hat{\theta} = \arg \left\{ \max_{\theta} \log L(\theta; \mathcal{Y}_{\mathcal{N}}) \right\}. \quad (11)$$

- In general, a numerical optimization method must be used

Uncertainty of ML estimates

- Approximate covariance of the parameter estimates

$$\text{Var} \left[\hat{\theta} \right] \simeq -H^{-1} \quad (12)$$

- Approximate Fisher information matrix

$$\{H\}_{ij} = \left. \frac{\partial^2 \log L(\theta; \mathcal{Y}_N)}{\partial \theta_i \partial \theta_j} \right|_{\theta=\hat{\theta}} \quad (13)$$

Maximum likelihood estimation (stationary)

ML estimates under stationary conditions

Likelihood under stationary conditions

$$L(\theta; \mathcal{Y}_N) = \prod_{t=1}^N [(2\pi)^m \det \Sigma]^{-\frac{1}{2}} \exp \left[-\frac{1}{2} \tilde{Y}_{t|t-1}^T \Sigma^{-1} \tilde{Y}_{t|t-1} \right] \quad (14)$$

$$= [(2\pi)^m \det \Sigma]^{N/2} \exp \left[-\frac{1}{2} \sum_{t=1}^N \tilde{Y}_{t|t-1}^T \Sigma^{-1} \tilde{Y}_{t|t-1} \right] \quad (15)$$

Stationarity implies that the matrices A , B , C , Σ_1 and Σ_2 are constant. Consequently, missing observations are not allowed

Σ known

Maximum likelihood estimation is equivalent to minimizing

$$S_1(\theta) = \frac{1}{N} \sum_{t=1}^N \tilde{Y}_{t|t-1}^T \Sigma^{-1} \tilde{Y}_{t|t-1} \quad (16)$$

$\tilde{Y}_{t+1|t} = Y_{t+1} - \hat{Y}_{t+1|t}$ is the one-step prediction error

Sample covariance of the predictions errors

$$D(\theta) = \frac{1}{N} \sum_{t=1}^N \tilde{Y}_{t|t-1} \tilde{Y}_{t|t-1}^T \quad (17)$$

Rewrite S_1

$$S_1(\theta) = \text{tr} \Sigma^{-1} D(\theta) \quad (18)$$

Σ unknown

- ▶ Maximum likelihood estimation is equivalent to minimizing

$$S(\theta, \Sigma) = \frac{1}{2}N \log \det \Sigma + \frac{1}{2} \sum_{t=1}^N \tilde{Y}_{t|t-1}^T \Sigma^{-1} \tilde{Y}_{t|t-1} \quad (19)$$

- ▶ Differentiate wrt. Σ

$$\frac{\partial S}{\partial \Sigma} = \frac{N}{2} \Sigma^{-1} - \frac{1}{2} \Sigma^{-1} \left(\sum_{t=1}^N \tilde{Y}_{t|t-1} \tilde{Y}_{t|t-1}^T \right) \Sigma^{-1} \quad (20)$$

Solution (ML estimator)

$$\Sigma = \frac{1}{N} \sum_{t=1}^N \tilde{Y}_{t|t-1} \tilde{Y}_{t|t-1}^T = D(\theta) \quad (21)$$

- ▶ For any given value of θ , (21) minimizes (19). The problem can then be reduced by the constraint $\Sigma = D(\theta)$.

Σ unknown

Substitute this constraint into the above cost function

$$S_c(\theta) = \frac{1}{2} \log \det D(\theta) + \frac{1}{2} \sum_{t=1}^N \tilde{Y}_{t|t-1}^T D^{-1}(\theta) \tilde{Y}_{t|t-1} \quad (22)$$

or

$$S_c(\theta) = \frac{1}{2} \log \det D(\theta) + \frac{1}{2} \text{tr} \left[\left(\sum_{t=1}^N \tilde{Y}_{t|t-1} \tilde{Y}_{t|t-1}^T \right) D^{-1}(\theta) \right] \quad (23)$$

$$= \frac{1}{2} \log \det D(\theta) + \frac{N}{2} \text{tr} I_m \quad (24)$$

$$= \frac{1}{2} \log \det D(\theta) + \frac{Nm}{2} \quad (25)$$

Since the last term is constant, the ML estimate is found by minimizing the *cost function*

$$S_2(\theta) = \log \det D(\theta) \quad (26)$$

Doubly stochastic processes

Estimation of some doubly stochastic models

- Consider the *non-linear* process:

$$Y_t = \Phi_t Y_{t-1} + \epsilon_t \quad (27)$$

$$\Phi_t - \mu = \phi(\Phi_{t-1} - \mu) + \zeta_t \quad (28)$$

where $\{\epsilon_t\}$ and $\{\zeta_t\}$ are mutually uncorrelated Gaussian white noise processes with variances σ_ϵ^2 and σ_ζ^2 , respectively.

- This model is some times referred to as an AR(1)-AR(1) model, since both the model for Y_t and the embedded model for the parameter variation is an AR(1) model.
- The parameters of the non-linear model are $(\phi, \mu, \sigma_\epsilon^2, \sigma_\zeta^2)$.

Estimation of some doubly stochastic models

Can be reparametrized as a state space model (more or less)

Rewrite model for parameter variation

$$\Phi_t = \phi \Phi_{t-1} + \mu(1 - \phi) + \zeta_t$$

Introduce $\delta = \mu(1 - \phi)$

$$\begin{pmatrix} \Phi_t \\ \delta_t \end{pmatrix} = \begin{pmatrix} \phi & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \Phi_{t-1} \\ \delta_{t-1} \end{pmatrix} + \begin{pmatrix} 1 \\ 0 \end{pmatrix} \zeta_t \quad (29)$$

$$Y_t = (Y_{t-1}, 0) \begin{pmatrix} \Phi_t \\ \delta_t \end{pmatrix} + \epsilon_t \quad (30)$$

Note: The new parameter is introduced as a state variable, and there is a unique relation between the new parameters $(\phi, \delta, \sigma_\epsilon^2, \sigma_\zeta^2)$ and the parameters of the model (27)

A generalization

The principle can be generalized

Consider, e.g., an AR(1)-AR(2) model

$$Y_t = \Phi_t Y_{t-1} + \epsilon_t \quad (31)$$

$$\Phi_t - \mu = \phi_1(\Phi_{t-1} - \mu) + \phi_2(\Phi_{t-2} - \mu) + \zeta_t \quad (32)$$

It can be written as the following linear state space model

$$\begin{pmatrix} \Phi_t \\ \Phi_{t-1} \\ \delta_t \end{pmatrix} = \begin{pmatrix} \phi_1 & \phi_2 & 1 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \Phi_{t-1} \\ \Phi_{t-2} \\ \delta_{t-1} \end{pmatrix} + \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \zeta_t \quad (33)$$

$$Y_t = (Y_{t-1}, 0, 0) \begin{pmatrix} \Phi_t \\ \Phi_{t-1} \\ \delta_t \end{pmatrix} + \epsilon_t \quad (34)$$

where $\delta = \mu(1 - \phi_1 - \phi_2)$.

Linear parameter-varying (LPV) systems

General form

$$x_{t+1} = A_t x_t + B_t u_t + v_t, \quad v_t \sim N(0, R_v), \quad (35a)$$

$$y_t = C_t x_t + e_t, \quad e_t \sim N(0, R_e) \quad (35b)$$

Examples of dependencies ($Y_{t-1} = [y_{t-1}, y_{t-2}, \dots]$)

$$C_t = C_t(Y_{t-1}), \quad (\text{deterministic function}) \quad (36a)$$

$$C_t \sim \mathcal{F}(Y_{t-1}) \quad (\text{probability distribution}) \quad (36b)$$

Measurement equation: Conditional distribution

Conditional mean

$$\begin{aligned}\hat{y}_{t|t-1} &= \mathbb{E}[y_t|Y_{t-1}] = \mathbb{E}[C_t x_t + e_t|Y_{t-1}] = C_t(Y_{t-1})\mathbb{E}[x_t|Y_{t-1}] \\ &= C_t(Y_{t-1})\hat{x}_{t|t-1}\end{aligned}\quad (37)$$

Conditional covariance

$$\begin{aligned}P_{t|t-1}^y &= \text{Cov}[y_t|Y_{t-1}] = \text{Cov}(C_t x_t + e_t|Y_{t-1}) \\ &= C_t(Y_{t-1})\text{Cov}(x_t|Y_{t-1})C_t^T(Y_{t-1}) + \text{Cov}(e_t|Y_{t-1}) \\ &= C_t(Y_{t-1})P_{t|t-1}C_t^T(Y_{t-1}) + R_e\end{aligned}\quad (38)$$

Conditional cross-covariance

$$\begin{aligned}P_{t|t-1}^{yx} &= \text{Cov}(y_t, x_t|Y_{t-1}) = \text{Cov}(C_t x_t + e_t, x_t|Y_{t-1}) \\ &= C_t(Y_{t-1})\text{Cov}(x_t, x_t|Y_{t-1}) + \text{Cov}(e_t, x_t|Y_{t-1}) \\ &= C_t(Y_{t-1})P_{t|t-1}\end{aligned}\quad (39)$$

Projection theorem

Joint distribution

$$\begin{bmatrix} x_t \\ y_t \end{bmatrix} \Big| Y_{t-1} \sim N \left(\begin{bmatrix} \hat{x}_{t|t-1} \\ C_t x_{t|t-1} \end{bmatrix}, \begin{bmatrix} P_{t|t-1} & P_{t|t-1} C_t^T \\ C_t P_{t|t-1} & C_t P_{t|t-1} C_t^T + R_e \end{bmatrix} \right), \quad (40a)$$

$$C_t = C_t(Y_{t-1}) \quad (40b)$$

Projection theorem

$$\begin{aligned} \mathbb{E}[x_t | y_t, Y_{t-1}] &= \mathbb{E}[x_t | Y_{t-1}] + K_t(y_t - C_t(Y_{t-1})\hat{x}_{t|t-1}), \\ &= \hat{x}_{t|t-1} + K_t(y_t - C_t(Y_{t-1})\hat{x}_{t|t-1}), \end{aligned} \quad (41a)$$

$$\text{Cov}(x_t | y_t, Y_{t-1}) = \text{Cov}(x_t | Y_{t-1}) - K_t C_t(Y_{t-1}) P_{t|t-1}, \quad (41b)$$

$$= P_{t|t-1} - K_t C_t(Y_{t-1}) P_{t|t-1} \quad (41c)$$

State equation: Conditional distribution

State equation: Conditional mean

$$\begin{aligned}\hat{x}_{t+1|t} &= \mathbb{E}[x_{t+1}|Y_t] = \mathbb{E}[A_t x_t + B_t u_t + v_t|Y_t] \\ &= A_t(Y_t)\mathbb{E}[x_t|Y_t] + B_t(Y_t)\mathbb{E}[u_t|Y_t] \\ &= A_t(Y_t)\hat{x}_{t|t} + B_t(Y_t)u_t\end{aligned}\tag{42}$$

State equation: Conditional covariance

$$\begin{aligned}P_{t+1|t} &= \text{Cov}(x_{t+1}|Y_t) = \text{Cov}(A_t x_t + B_t u_t + v_t|Y_t) \\ &= A_t(Y_t)\text{Cov}(x_t|Y_t)A_t^T(Y_t) + \text{Cov}(v_t|Y_t) \\ &= A_t(Y_t)P_{t|t}A_t^T(Y_t) + R_v\end{aligned}\tag{43}$$

Kalman filter for a class of LPV systems

Measurement update / data update / filtering step

$$\hat{x}_{t|t} = \hat{x}_{t|t-1} + K_t(y_t - C_t(Y_{t-1})\hat{x}_{t|t-1}), \quad (44a)$$

$$P_{t|t} = P_{t|t-1} - K_t C_t(Y_{t-1}) P_{t|t-1} \quad (44b)$$

Time update / prediction step

$$\hat{x}_{t+1|t} = A_t(Y_t)\hat{x}_{t|t} + B_t(Y_t)u_t, \quad (45a)$$

$$P_{t+1|t} = A_t(Y_t)P_{t|t}A_t^T(Y_t) + R_v \quad (45b)$$

Extended Kalman filter

The non-linear, time-varying, discrete time state space model

Nonlinear discrete-time system

$$X_{t+1} = f(t, X_t, U_t) + v_t \quad (46)$$

$$Y_t = h(t, X_t, U_t) + e_t \quad (47)$$

$\{e_t\}$ and $\{v_t\}$ are sequences of independent random variables

$$\mathbb{E}[vv^T] = R_{1,t}$$

$$\mathbb{E}[ee^T] = R_{2,t}$$

$$\mathbb{E}[ve^T] = 0$$

The Extended Kalman Filter

Extended Kalman filter (EKF): Estimate of X_{t+1} given $Y_0, Y_1, \dots, Y_t$

$$\hat{X}_{t+1|t} = f(t, \hat{X}_{t|t-1}, U_t) + K_t[Y_t - h(t, \hat{X}_{t|t-1})] \quad (48)$$

Kalman gain

$$K_t = F_t P_{t|t-1} H_t^T (H_t P_{t|t-1} H_t^T + R_{2,t})^{-1}$$

Covariance

$$P_{t+1|t} = F_t P_{t|t-1} F_t^T + R_{1,t} - K_t (H_t P_{t|t-1} H_t^T + R_{2,t}) K_t^T$$

Jacobians of f and h

$$\begin{aligned} F_t &= \frac{\partial f}{\partial X}(\hat{X}_{t|t-1}, U_t), \\ H_t &= \frac{\partial h}{\partial X}(\hat{X}_{t|t-1}, U_t) \end{aligned}$$

Extended Kalman filter for parameter estimation

Extended Kalman Filter for parameter estimation

Main idea: Include the unknown parameters in the state vector and then use a state estimation method to estimate both the state and the parameters

Important: The resulting state space model is in general a non-linear state space model, and the *extended Kalman filter* is then often used

Linear state space model

$$\begin{aligned}X_{t+1} &= A(\theta)X_t + B(\theta)u_t + v_t \\Y_t &= C(\theta)X_t + e_t\end{aligned}$$

the parameter vector θ is unknown

Extended Kalman Filter for parameter estimation

By including $(\theta = \theta_t)$ in a new state vector

$$Z_t = \begin{pmatrix} X_t \\ \theta_t \end{pmatrix}$$

this leads to the following non-linear state space model

$$Z_{t+1} = f(Z_t, u_t) + \begin{pmatrix} v_t \\ 0 \end{pmatrix}$$

$$Y_t = h(Z_t) + e_t$$

where

$$f(Z_t, u_t) = \begin{pmatrix} A(\theta)X_t + B(\theta)u_t \\ \theta_t \end{pmatrix}$$

$$h(Z_t) = C(\theta)X_t$$

EKF state estimation will simultaneously estimate the original state vector and the parameter vector θ

Generalized state space models

Generalized State Space Models

- ▶ Sequence of state variables

$$\{X_t; t \geq 1\}$$

- ▶ Sequence of observations

$$\{Y_t; t \geq 1\}$$

- ▶ Only the one-dimensional case is considered here

Parameter-driven models

The model evolves independently of the past history of the observation process

Let the history of $\{X_t\}$ and $\{Y_t\}$ be denoted by $\mathcal{X}^{(t-1)}$ and $\mathcal{Y}^{(t-1)}$:

$$\mathcal{X}^{(t-1)} = (X_{t-1}, X_{t-2}, \dots)$$

$$\mathcal{Y}^{(t-1)} = (Y_{t-1}, Y_{t-2}, \dots)$$

Assumption: Y_t given $(X_t, \mathcal{X}^{(t-1)}, \mathcal{Y}^{(t-1)})$ is independent of $(\mathcal{X}^{(t-1)}, \mathcal{Y}^{(t-1)})$ with conditional probability density:

$$p(y_t | x_t) := p(y_t \mid x_t, \mathcal{X}^{(t-1)}, \mathcal{Y}^{(t-1)}) \quad (\text{obs. eq.})$$

And similarly

$$p(x_{t+1} | x_t) := p(x_{t+1} \mid x_t, \mathcal{X}^{(t-1)}, \mathcal{Y}^{(t-1)}) \quad (\text{state eq.})$$

Joint density of observation and state variables

$$\begin{aligned} p(y_1, \dots, y_n, x_1, \dots, x_n) &= p(y_n | x_n, \mathcal{X}^{(n-1)}, \mathcal{Y}^{(n-1)}) p(x_n, \mathcal{X}^{(n-1)}, \mathcal{Y}^{(n-1)}) \\ &= p(y_n | x_n) p(x_n | \mathcal{X}^{(n-1)}, \mathcal{Y}^{(n-1)}) p(\mathcal{Y}^{(n-1)}, \mathcal{X}^{(n-1)}) \\ &= \dots \\ &= \prod_{j=1}^n p(y_j | x_j) \left(\prod_{j=2}^n p(x_j | x_{j-1}) \right) p(x_1) \end{aligned}$$

where $p(x_1)$ denotes the probability density of the initial state.

- ▶ $\{X_t\}$ is a first order Markov process.
- ▶ It is also clearly seen that

$$p(y_1, \dots, y_n | x_1, \dots, x_n) = \prod_{j=1}^n p(y_j | x_j)$$

- ▶ that is, the dependence structure of $\{Y_t\}$ is inherited from that of $\{X_t\}$.
- ▶ The sequence $\{X_t\}$ is often referred to as the hidden or latent process. In conclusion we have a **Hidden Markov Process**.

Filtering and prediction

We want to determine $p(x_t|\mathcal{Y}^{(t)})$ and $p(x_{t+1}|\mathcal{Y}^{(t)})$

Bayes' theorem

$$p(x_t|\mathcal{Y}^{(t)}) = \frac{p(y_t|x_t) p(x_t|\mathcal{Y}^{(t-1)})}{p(y_t|\mathcal{Y}^{(t-1)})}$$

and

$$p(x_{t+1}|\mathcal{Y}^{(t)}) = \int p(x_t|\mathcal{Y}^{(t)}) p(x_{t+1}|x_t) dx_t$$

In the case of a discrete state space, a summation is used instead of the integral

Prediction of future observations

$$p(y_{t+1}|\mathcal{Y}^{(t)}) = \int p(y_{t+1}|x_{t+1}) p(x_{t+1}|\mathcal{Y}^{(t)}) dx_{t+1}$$

In

$$p(y_1, \dots, y_n, x_1, \dots, x_n) = \prod_{j=1}^n p(y_j|x_j) \left(\prod_{j=2}^n p(x_j|x_{j-1}) \right) p(x_1)$$

the left side is the a posteriori distribution, and the right side is a multiplication of the likelihood and the prior.

A non-Gaussian example

- ▶ In general, the recursions can be very difficult to compute
- ▶ Consider a very simple example, with the state equation

$$X_{t+1} = aX_t \quad (*)$$

- ▶ The *observation density* is assumed to be

$$p(y_t|x_t) = \frac{(\pi x_t)^{y_t} e^{-\pi x_t}}{y_t!}, \quad y_t = 0, 1, 2, \dots$$

where $\pi \in]0, 1]$ is constant.

- ▶ (*) implies that (δ is Dirac's delta)

$$p(x_{t+1}|x_t) = \delta(x_{t+1} - ax_t)$$

- ▶ Assume further that X_1 is *gamma-distributed*:

$$p_1(x_1) = g(x_1; \alpha, \lambda) = \frac{\lambda^\alpha x_1^{\alpha-1} e^{-\lambda x_1}}{\Gamma(\alpha)} \quad (x_1 > 0)$$

- This could be a simple model for the *evolution of the number, X_t , of individuals* at time t *infected with a rare disease* in which X_t is treated as a continuous variable.
- The *observation Y_t represents the number of infected individuals observed in a random sample consisting of a small fraction π of the population a time, t .*
- Now

$$\begin{aligned}
 p(x_1|y_1) &= p(y_1|x_1) p(x_1) / p(y_1) \\
 &= \left(\frac{(\pi x_1)^{y_1} e^{-\pi x_1}}{y_1!} \right) \left(\frac{\lambda^\alpha x_1^{\alpha-1} e^{-\lambda x_1}}{\Gamma(\alpha)} \right) \frac{1}{p(y_1)} \\
 &= c(y_1) x_1^{\alpha+y_1-1} e^{-(\pi+\lambda)x_1} \quad (x_1 > 0) \\
 &= g(x_1; \underbrace{\alpha_1}_{\alpha+y_1}, \underbrace{\lambda_1}_{\pi+\lambda}) \quad (\text{again gamma distributed})
 \end{aligned}$$

- The prediction density becomes

$$p(x_2|\mathcal{Y}^{(1)}) = g(x_2; \alpha_1, \lambda_1/a)$$

Iterating the recursions, we find that

$$p(x_t | \mathcal{Y}^{(t)}) = g(x_t; \alpha_t, \lambda_t)$$

and

$$p(x_{t+1} | \mathcal{Y}^{(t)}) = g(x_{t+1}; \alpha_t, \lambda_t/a)$$

where

$$\alpha_t = \alpha_{t-1} + y_t = \alpha + y_1 + \dots + y_t$$

$$\lambda_t = \lambda_{t-1}/a + \pi = \lambda a^{1-t} + \pi \frac{(1 - a^{-t})}{(1 - a^{-1})}$$

Observation-driven models

Assumption: (again) Y_t given $(X_t, \mathcal{X}^{(t-1)}, \mathcal{Y}^{(t-1)})$ is independent of $(\mathcal{X}^{(t-1)}, \mathcal{Y}^{(t-1)})$.

The model is specified by

$$p(y_t|x_t) = p(y_t|\mathcal{X}^{(t)}, \mathcal{Y}^{(t-1)}) \quad (\text{obs. eqn.})$$

$$p(x_{t+1}|\mathcal{Y}^{(t)}) = \text{depends on all obs.} \quad (\text{state eqn.})$$

The advantage of the observation-driven state equation is that Bayes' Theorem can be used directly without the use of the updating formula.

Note: The observation equation is the same for both types of models but the state vector of a parameter-driven model evolves independent of past observation, while the state vector of an observation-driven model depends on past observations.

Example – observation-driven model

An AR(1) process with i.i.d noise can be expressed as an observation-driven model.

Suppose

$$Y_t = \Phi Y_{t-1} + Z_t$$

$\{Z_t\}$ is i.i.d with mean 0 and probability density $f(x)$. Then with

$$X_t := Y_{t-1}$$

we have

$$\begin{aligned} p(y_t | x_t) &= f(y_t - \Phi x_t) \\ p(x_{t+1} | \mathcal{Y}^{(t)}) &= \delta(x_{t+1} - y_t) \end{aligned}$$