

ADVANCED MACHINE LEARNING

STATE-SPACE MODELS

LECTURE 1

Mattias Villani

Division of Statistics and Machine Learning
Department of Computer and Information Science
Linköping University



LECTURE OVERVIEW

- ▶ Time varying parameter models
- ▶ State space models
- ▶ The Bayes filter
- ▶ The Kalman filter

AUTOREGRESSIVE TIME SERIES MODELS

- ▶ Autoregressive process (AR) for time series

$$y_t = \rho y_{t-1} + \varepsilon_t, \quad \varepsilon_t \stackrel{iid}{\sim} N(0, \sigma^2)$$

- ▶ The joint distribution for the whole time sequence y_1, y_2, \dots, y_T factorizes as

$$p(y_1, \dots, y_T) = p(y_1)p(y_2|y_1) \cdots p(y_T|y_{T-1})$$

where

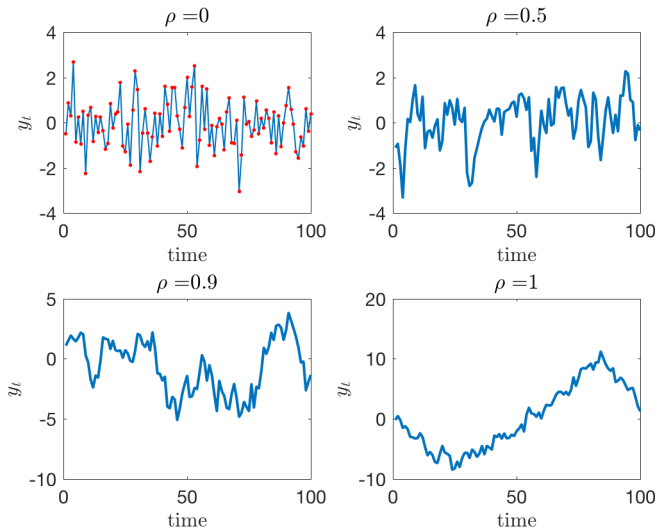
$$y_t|y_{t-1} \sim N(\rho y_{t-1}, \sigma^2).$$

- ▶ $AR(p)$ process

$$y_t|y_{t-1}, \dots, y_{t-p} \sim N\left(\sum_{j=1}^p \rho_j y_{t-j}, \sigma^2\right).$$

- ▶ $ARIMA(p, q)$.

AUTOREGRESSIVE TIME SERIES MODELS



HIDDEN MARKOV MODELS

- ▶ Two **regimes** defined by **latent** (**hidden**) variable $x_t \in \{1, 2\}$

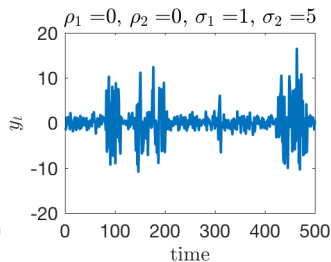
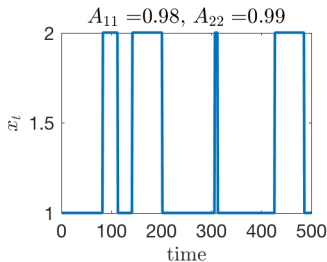
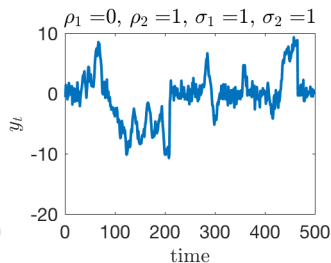
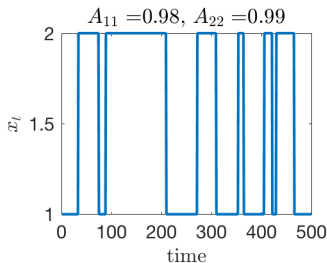
$$y_t = \begin{cases} \rho_1 y_{t-1} + \varepsilon_t, & \varepsilon_t \stackrel{iid}{\sim} N(0, \sigma_1^2) & \text{if } z_t = 1 \\ \rho_2 y_{t-1} + \varepsilon_t, & \varepsilon_t \stackrel{iid}{\sim} N(0, \sigma_2^2) & \text{if } z_t = 2 \end{cases}$$

- ▶ x_t follows a **Markov chain**. Transition from state $j \rightarrow k$

$$\Pr(x_t = k | x_{t-1} = j) = A_{jk}$$

- ▶ But what if changes in parameters appear **more gradual**?

HIDDEN MARKOV MODELS



TIME VARYING PARAMETER MODELS

- ▶ Smoothly **time varying parameter model**

$$y_t = \rho_t y_{t-1} + \varepsilon_t \quad \varepsilon_t \stackrel{iid}{\sim} N(0, \sigma_\varepsilon^2)$$

$$\rho_t = \rho_{t-1} + v_t \quad v_t \stackrel{iid}{\sim} N(0, \sigma_v^2)$$

- ▶ The persistence parameter ρ is a **latent (hidden) continuous variable** that evolves over time (random walk).
- ▶ More generally, for some $-1 \leq a < 1$,

$$y_t = \rho_t y_{t-1} + \varepsilon_t \quad \varepsilon_t \stackrel{iid}{\sim} N(0, \sigma_\varepsilon^2)$$

$$\rho_t = a\rho_{t-1} + v_t \quad v_t \stackrel{iid}{\sim} N(0, \sigma_v^2)$$

- ▶ **Time varying variance**

$$y_t = \rho y_{t-1} + \varepsilon_t \quad \varepsilon_t \stackrel{iid}{\sim} N(0, \sigma_{\varepsilon,t}^2)$$

$$\ln \sigma_{\varepsilon,t}^2 = \ln \sigma_{\varepsilon,t-1}^2 + v_t \quad v_t \stackrel{iid}{\sim} N(0, \sigma_v^2)$$

TIME VARYING PARAMETER MODELS

- ▶ Smoothly **time varying parameter regression**

$$y_t = \mathbf{x}_t^T \beta_t + \varepsilon_t \quad \varepsilon_t \stackrel{iid}{\sim} N(0, \sigma_\varepsilon^2)$$

$$\beta_t = \beta_{t-1} + v_t \quad v_t \stackrel{iid}{\sim} N(0, \sigma_v^2)$$

- ▶ Smoothly **time varying parameter survival model**
- ▶ The **hazard function** (conditional probability of death at time t):

$$\lambda(t|\mathbf{x}) = \lambda_0(t) \cdot \exp(\mathbf{x}^T \beta_t)$$

$$\beta_t = \beta_{t-1} + v_t \quad v_t \stackrel{iid}{\sim} N(0, \sigma_v^2)$$

- ▶ **Locally linear trend model**

$$y_t = \mu_t + \varepsilon_t \quad \varepsilon_t \stackrel{iid}{\sim} N(0, \sigma_\varepsilon^2)$$

$$\mu_t = \mu_{t-1} + \beta_t + v_t \quad v_t \stackrel{iid}{\sim} N(0, \sigma_v^2)$$

$$\beta_t = \beta_{t-1} + \eta_t \quad \eta_t \stackrel{iid}{\sim} N(0, \sigma_\eta^2)$$

STATE-SPACE MODELS

- ▶ Basic **state-space model**

$$\text{Measurement eq: } y_t = Cx_t + \varepsilon_t \quad \varepsilon_t \stackrel{iid}{\sim} N(0, \sigma_\varepsilon^2)$$

$$\text{State eq: } x_t = Ax_{t-1} + v_t \quad v_t \stackrel{iid}{\sim} N(0, \sigma_v^2)$$

- ▶ **Measurements** y_t are driven by an underlying unobserved **state** x_t .
- ▶ **Time-varying parameter models**: $x_t = \rho_t$.
- ▶ **Hidden Markov models** are state space models with a **discrete state** variable.
- ▶ Example 1: x_t is employment at time t . y_t are labor force survey estimates.
- ▶ Example 2: x_t is democrats' voting share. y_t are results from poll.
- ▶ Example 3: x_t is the position of flying vehicle at time t . y_t are sensor measurements.

STATE SPACE MODELS

- ▶ The **linear Gaussian state-space (LGSS) model**

$$\text{Measurement eq: } \mathbf{y}_t = \mathbf{C}\mathbf{x}_t + \varepsilon_t \quad \varepsilon_t \stackrel{iid}{\sim} N(0, \Omega_\varepsilon)$$

$$\text{State eq: } \mathbf{x}_t = \mathbf{A}\mathbf{x}_{t-1} + \mathbf{B}\mathbf{u}_t + \nu_t \quad \nu_t \stackrel{iid}{\sim} N(0, \Omega_\nu)$$

- ▶ \mathbf{u}_t is a vector of **control variables** that affect the state.
- ▶ Example 1: In robotics, \mathbf{u}_t are control commands (steering, gas, brake etc).
- ▶ Example 2: In economics, \mathbf{u}_t could be the central banks increase/decrease of the interest rate.
- ▶ Example 3: \mathbf{u}_t could be amount spent on political campaigns.

THE POSTERIOR DISTRIBUTION OF THE STATE

- ▶ The **linear Gaussian state-space (LGSS)** model

$$\text{Measurement eq: } \mathbf{y}_t = \mathbf{C}\mathbf{x}_t + \varepsilon_t \quad \varepsilon_t \stackrel{iid}{\sim} N(0, \Omega_\varepsilon)$$

$$\text{State eq: } \mathbf{x}_t = \mathbf{A}\mathbf{x}_{t-1} + \mathbf{B}\mathbf{u}_t + \nu_t \quad \nu_t \stackrel{iid}{\sim} N(0, \Omega_\nu)$$

- ▶ Aim: the **posterior distribution of the state** at time t

$$p(\mathbf{x}_t | \mathbf{y}_1, \dots, \mathbf{y}_T, \mathbf{u}_1, \dots, \mathbf{u}_T)$$

- ▶ Also called the **smoothing distribution**.
- ▶ The **joint smoothing distribution**

$$p(\mathbf{x}_1, \dots, \mathbf{x}_T | \mathbf{y}_1, \dots, \mathbf{y}_T, \mathbf{u}_1, \dots, \mathbf{u}_T)$$

- ▶ More on this later.

MODEL STRUCTURE

- ▶ The **linear Gaussian state-space (LGSS)** model

$$\text{Measurement eq: } \mathbf{y}_t = \mathbf{C}\mathbf{x}_t + \varepsilon_t \quad \varepsilon_t \stackrel{iid}{\sim} N(0, \Omega_\varepsilon)$$

$$\text{State eq: } \mathbf{x}_t = \mathbf{A}\mathbf{x}_{t-1} + \mathbf{B}\mathbf{u}_t + \nu_t \quad \nu_t \stackrel{iid}{\sim} N(0, \Omega_\nu)$$

- ▶ Note 1: \mathbf{x}_t is first order Markov: $p(\mathbf{x}_t | \mathbf{x}_{t-1}, \dots, \mathbf{x}_1) = p(\mathbf{x}_t | \mathbf{x}_{t-1})$.
- ▶ Note 2: Conditional on \mathbf{x}_t , \mathbf{y}_t is independent of past observations and states.
- ▶ State space as **graphical model**.

THE FILTERING DISTRIBUTION

- ▶ Short hand notation: $\mathbf{x}_{1:t} = \{\mathbf{x}_1, \dots, \mathbf{x}_t\}$.
- ▶ Aim: the **filtering distribution of the state** at time t

$$p(\mathbf{x}_t | \mathbf{y}_{1:t}, \mathbf{u}_{1:t})$$

- ▶ Short hand for the **posterior** (belief) for \mathbf{x}_t

$$\text{bel}(\mathbf{x}_t) \equiv p(\mathbf{x}_t | \mathbf{y}_{1:t}, \mathbf{u}_{1:t})$$

- ▶ Short hand for the **prior** (belief) for \mathbf{x}_t , before the measurement at time t ,

$$\overline{\text{bel}}(\mathbf{x}_t) \equiv p(\mathbf{x}_t | \mathbf{y}_{1:t-1}, \mathbf{u}_{1:t})$$

THE BAYES FILTER

- ▶ We are now at time t .
- ▶ We have just given the control command \mathbf{u}_t .
- ▶ We have not yet observed \mathbf{y}_t .
- ▶ Our beliefs at this stage:

$$\overline{\text{bel}}(\mathbf{x}_t) = \int p(\mathbf{x}_t | \mathbf{u}_t, \mathbf{x}_{t-1}) \text{bel}(\mathbf{x}_{t-1}) d\mathbf{x}_{t-1}$$

- ▶ Now comes the observation \mathbf{y}_t .
- ▶ **Update your beliefs** using Bayes' theorem:

$$\text{bel}(\mathbf{x}_t) \propto p(\mathbf{y}_t | \mathbf{x}_t) \overline{\text{bel}}(\mathbf{x}_t).$$

THE BAYES FILTER

- **Prediction step** (control update)

$$\overline{\text{bel}}(\mathbf{x}_t) = \int p(\mathbf{x}_t | \mathbf{u}_t, \mathbf{x}_{t-1}) \text{bel}(\mathbf{x}_{t-1}) d\mathbf{x}_{t-1}$$

- **Measurement update step**

$$\text{bel}(\mathbf{x}_t) \propto p(\mathbf{y}_t | \mathbf{x}_t) \overline{\text{bel}}(\mathbf{x}_t).$$

THE KALMAN FILTER

- ▶ The **Kalman filter** is the special case of the Bayes filter for the linear Gaussian state-space (LGSS) model.
- ▶ Under **linearity** and **Gaussianity**:
 - ▶ we can compute the integral in the prediction step analytically
 - ▶ the posterior in the measurement update becomes Gaussian

- ▶ **Prediction update**

$$\overline{\text{bel}}(\mathbf{x}_t) = N(\bar{\mu}_t, \bar{\Sigma}_t)$$

- ▶ **Measurement update**

$$\text{bel}(\mathbf{x}_t) = N(\mu_t, \Sigma_t)$$

- ▶ The Kalman filter tells us how to **iteratively** compute the sequences $\{\mu_t, \Sigma_t\}$ throughout time $t = 1, \dots, T$.

THE KALMAN FILTER

- ▶ The **linear Gaussian state-space (LGSS)** model

Measurement eq: $\mathbf{y}_t = \mathbf{C}\mathbf{x}_t + \varepsilon_t \quad \varepsilon_t \stackrel{iid}{\sim} N(0, \Omega_\varepsilon)$

State eq: $\mathbf{x}_t = \mathbf{A}\mathbf{x}_{t-1} + \mathbf{B}\mathbf{u}_t + \nu_t \quad \nu_t \stackrel{iid}{\sim} N(0, \Omega_\nu)$

- ▶ **Algorithm KalmanFilter**($\mu_{t-1}, \Sigma_{t-1}, \mathbf{u}_t, \mathbf{y}_t$)

- ▶ Prediction update:
$$\begin{cases} \bar{\mu}_t = \mathbf{A}\mu_{t-1} + \mathbf{B}\mathbf{u}_t \\ \bar{\Sigma}_t = \mathbf{A}\Sigma_{t-1}\mathbf{A}^T + \Omega_\nu \end{cases}$$
- ▶ Measurement update :
$$\begin{cases} \mathbf{K}_t = \bar{\Sigma}_t \mathbf{C}^T (\mathbf{C}\bar{\Sigma}_t \mathbf{C}^T + \Omega_\varepsilon)^{-1} \\ \mu_t = \bar{\mu}_t + \mathbf{K}_t(\mathbf{y}_t - \mathbf{C}\bar{\mu}_t) \\ \Sigma_t = (\mathbf{I} - \mathbf{K}_t \mathbf{C}) \bar{\Sigma}_t \end{cases}$$
- ▶ Return μ_t, Σ_t

KALMAR FILTER INTUITION

- Assume everything is univariate and no control:

$$\text{Measurement eq: } y_t = cx_t + \varepsilon_t \quad \varepsilon_t \stackrel{iid}{\sim} N(0, \omega_\varepsilon^2)$$

$$\text{State eq: } x_t = ax_{t-1} + v_t \quad v_t \stackrel{iid}{\sim} N(0, \omega_v^2)$$

- **Algorithm KalmanFilter**($\mu_{t-1}, \sigma_{t-1}^2, y_t$)

- Prediction update:
$$\begin{cases} \bar{\mu}_t = a\mu_{t-1} \\ \bar{\sigma}_t = a^2\sigma_{t-1}^2 + \omega_v^2 \end{cases}$$

- Measurement update :
$$\begin{cases} k_t = \frac{c\bar{\sigma}_t^2}{c^2\bar{\sigma}_t^2 + \omega_\varepsilon^2} \\ \mu_t = \bar{\mu}_t + k_t(y_t - c\bar{\mu}_t) \\ \Sigma_t = \left(\frac{\omega_\varepsilon^2}{c^2\bar{\sigma}_t^2 + \omega_\varepsilon^2} \right) \bar{\sigma}_t^2 \end{cases}$$

A SIMULATED EXAMPLE

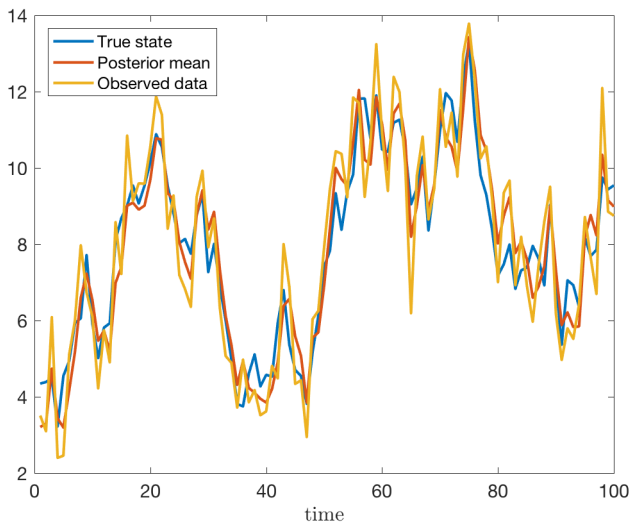
- ▶ The **linear Gaussian state-space (LGSS)** model

$$\text{Measurement eq: } y_t = x_t + \varepsilon_t \quad \varepsilon_t \stackrel{iid}{\sim} N(0, 1)$$

$$\text{State eq: } x_t = 0.9x_{t-1} + u_t + v_t \quad v_t \stackrel{iid}{\sim} N(0, 0.5)$$

- ▶ Control: $u_t \sim |r_t|$ where $r_t \sim N(0, 1)$.
- ▶ $T = 100$.
- ▶ Initial state value: $x_0 \sim N(0, 10^2)$.

DATA, STATE AND POSTERIOR OF STATE



POSTERIOR INTERVALS FOR THE STATE

