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**

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

- The **joint distribution** for the whole time sequence $y_1, y_2, ..., y_T$ factorizes as

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

where

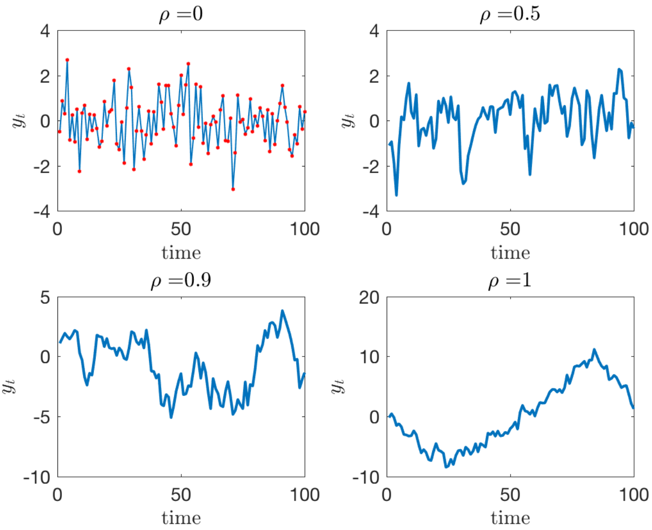
$$y_t|y_{t-1} \sim N\left(\rho y_{t-1}, \sigma^2\right).$$

- $AR(p)$ process

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

- $ARIMA(p, q)$.

Autoregressive time series models



Hidden Markov models

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

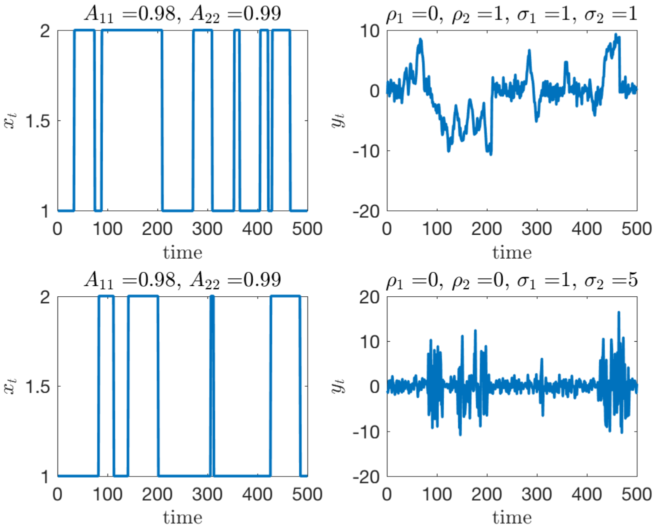
$$y_t = \begin{cases} \rho_1 y_{t-1} + \varepsilon_t, & \varepsilon_t \overset{iid}{\sim} N(0, \sigma_1^2) \quad \text{if } z_t = 1 \\ \rho_2 y_{t-1} + \varepsilon_t, & \varepsilon_t \overset{iid}{\sim} N(0, \sigma_2^2) \quad \text{if } z_t = 2 \end{cases}$$

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

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

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

Hidden Markov models



Time varying parameter models

- Smoothly **time varying parameter model**

$$y_t = \rho_t y_{t-1} + \varepsilon_t \qquad \varepsilon_t \overset{iid}{\sim} N\left(0, \sigma_\varepsilon^2\right)$$
$$\rho_t = \rho_{t-1} + \nu_t \qquad \nu_t \overset{iid}{\sim} N\left(0, \sigma_\nu^2\right)$$

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

$$y_t = \rho_t y_{t-1} + \varepsilon_t \qquad \varepsilon_t \overset{iid}{\sim} N\left(0, \sigma_\varepsilon^2\right)$$
$$\rho_t = a\rho_{t-1} + \nu_t \qquad \nu_t \overset{iid}{\sim} N\left(0, \sigma_\nu^2\right)$$

- **Time varying variance**

$$y_t = \rho y_{t-1} + \varepsilon_t \qquad \varepsilon_t \overset{iid}{\sim} N\left(0, \sigma_{\varepsilon,t}^2\right)$$
$$\ln \sigma_{\varepsilon,t}^2 = \ln \sigma_{\varepsilon,t-1}^2 + \nu_t \qquad \nu_t \overset{iid}{\sim} N\left(0, \sigma_\nu^2\right)$$

Time varying parameter models

- Smoothly **time varying parameter regression**

$$y_t = \mathbf{x}_t^T \beta_t + \varepsilon_t \qquad \varepsilon_t \overset{iid}{\sim} N\left(0, \sigma_\varepsilon^2\right)$$
$$\beta_t = \beta_{t-1} + \nu_t \qquad \nu_t \overset{iid}{\sim} N\left(0, \sigma_\nu^2\right)$$

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

$$\lambda(t|\mathbf{x}) = \lambda_0(t) \cdot \exp\left(\mathbf{x}^T \beta_t\right)$$
$$\beta_t = \beta_{t-1} + \nu_t \qquad \nu_t \overset{iid}{\sim} N\left(0, \sigma_\nu^2\right)$$

- And so on ...

Unobserved components models

- Model a time series as components: mean, trend, season, cycles etc.
- **Local level model**

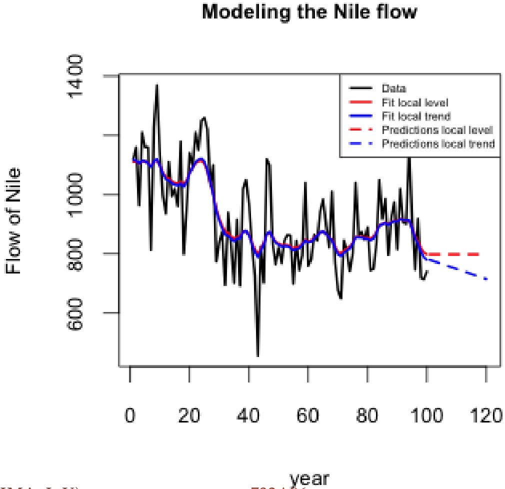
$$y_t = \mu_t + \varepsilon_t \qquad \varepsilon_t \overset{iid}{\sim} N\left(0, \sigma_\varepsilon^2\right)$$
$$\mu_t = \mu_{t-1} + \nu_t \qquad \nu_t \overset{iid}{\sim} N\left(0, \sigma_\nu^2\right)$$

- **Local trend model**

$$y_t = \mu_t + \varepsilon_t \qquad \varepsilon_t \overset{iid}{\sim} N\left(0, \sigma_\varepsilon^2\right)$$
$$\mu_t = \mu_{t-1} + \beta_{t-1} + \nu_t \qquad \nu_t \overset{iid}{\sim} N\left(0, \sigma_\nu^2\right)$$
$$\beta_t = \beta_{t-1} + \eta_t \qquad \eta_t \overset{iid}{\sim} N\left(0, \sigma_\eta^2\right)$$

Unobserved components models

- See my code `UnobservedComponentsModel.R`



State-space models

- Basic **state-space model**

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

$$\text{State eq:} \quad x_t = Ax_{t-1} + \nu_t \qquad \nu_t \overset{iid}{\sim} N\left(0, \sigma_\nu^2\right)$$

- **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: $\mathbf{x}_t$ is the position of flying vehicle at time t. $\mathbf{y}_t$ are sensor measurements.

Local trend model is a state space model

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

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

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

- **Local trend model**

$$y_t = \mu_t + \varepsilon_t \qquad \varepsilon_t \overset{iid}{\sim} N\left(0, \sigma_\varepsilon^2\right)$$

$$\mu_t = \mu_{t-1} + \beta_{t-1} + \nu_t \qquad \nu_t \overset{iid}{\sim} N\left(0, \sigma_\nu^2\right)$$

$$\beta_t = \beta_{t-1} + \eta_t \qquad \eta_t \overset{iid}{\sim} N\left(0, \sigma_\eta^2\right)$$

- State space formulation

$$x_t = \begin{pmatrix} \mu_t \\ \beta_t \end{pmatrix}, \mathbf{A} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \mathbf{C} = \begin{pmatrix} 1 & 0 \end{pmatrix}, \Omega_\varepsilon = \sigma_\varepsilon^2, \Omega_\nu = \begin{pmatrix} \sigma_{\nu 1}^2 & 0 \\ 0 & \sigma_{\nu 2}^2 \end{pmatrix}$$

The posterior distribution of the state

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

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

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

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

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

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

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

- More on this later.

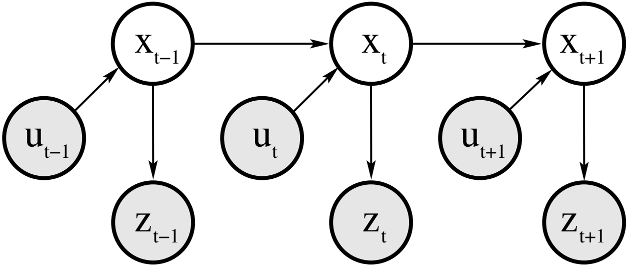
Model structure

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

$$
\begin{aligned}
\text{Measurement eq:}\quad \mathbf{y}_t &= \mathbf{C}\mathbf{x}_t + \varepsilon_t \qquad \varepsilon_t \overset{iid}{\sim} N\left(0, \Omega_\varepsilon\right) \\
\text{State eq:}\quad \mathbf{x}_t &= \mathbf{A}\mathbf{x}_{t-1} + \mathbf{B}\mathbf{u}_t + \nu_t \qquad \nu_t \overset{iid}{\sim} N\left(0, \Omega_\nu\right)
\end{aligned}
$$

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

Model structure



The filtering distribution

- Short hand notation: $\mathbf{x}_{1:t} = \{\mathbf{x}_1, ..., \mathbf{x}_t\}$.
- 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 $\mathbf{x}_t$

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

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

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

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:

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

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

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

The Bayes filter

- **Prediction step** (control update)

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

- **Measurement update step**

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

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**

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

- **Measurement update**

$$\mathrm{bel}(\mathbf{x}_t) = N\left(\mu_t, \Sigma_t\right)$$

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

The Kalman filter

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

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

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

- **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 : $\begin{cases} \mathbf{K}_t = \bar{\Sigma}_t\mathbf{C}^T\left(\mathbf{C}\bar{\Sigma}_t\mathbf{C}^T + \Omega_\varepsilon\right)^{-1} \\ \mu_t = \bar{\mu}_t + \mathbf{K}_t(\mathbf{y}_t - \mathbf{C}\bar{\mu}_t) \\ \Sigma_t = (\mathbf{I} - \mathbf{K}_t\mathbf{C})\bar{\Sigma}_t \end{cases}$
  - Return μ_t, Σ_t

Kalmar filter intuition

- Assume everything is univariate and no control:

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

$$\text{State eq:}\quad x_t = ax_{t-1} + \nu_t \qquad \nu_t \overset{iid}{\sim} N\left(0, \omega_\nu^2\right)$$

- **Algorithm KalmanFilter**$(\mu_{t-1}, \sigma_{t-1}^2, y_t)$
- Prediction update: $\begin{cases} \bar{\mu}_t = a\mu_{t-1} \\ \bar{\sigma}_t = a^2\sigma_{t-1}^2 + \omega_\nu^2 \end{cases}$
- Measurement update : $\begin{cases} k_t = \frac{c\bar{\sigma}_t^2}{c^2\bar{\sigma}_t^2 + \omega_\varepsilon^2} \\ \mu_t = \bar{\mu}_t + k_t(y_t - c\bar{\mu}_t) \\ \Sigma_t = \left(\frac{\omega_\varepsilon^2}{c^2\bar{\sigma}_t^2 + \omega_\varepsilon^2}\right)\bar{\sigma}_t^2 \end{cases}$

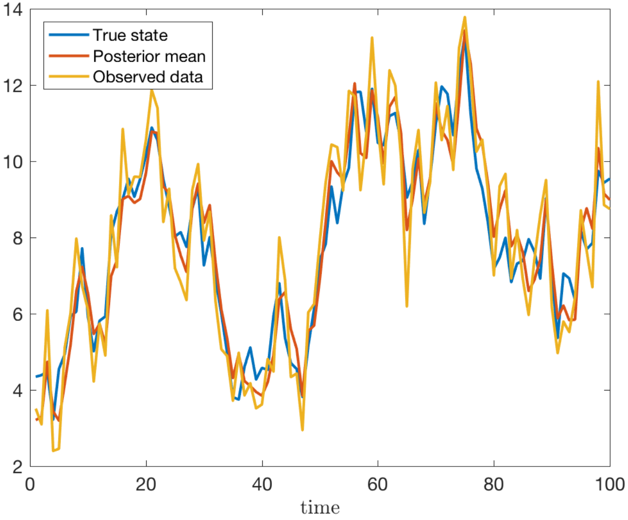
A simulated example

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

$$\text{Measurement eq:} \quad y_t = x_t + \varepsilon_t \qquad \varepsilon_t \overset{iid}{\sim} N(0,1)$$
$$\text{State eq:} \quad x_t = 0.9x_{t-1} + u_t + \nu_t \qquad \nu_t \overset{iid}{\sim} N(0,0.5)$$

- 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