# State-Space Models

## Models, Applications and State inference

**Mattias Villani**

Statistiska institutionen
Stockholms universitet

Institutionen för datavetenskap
Linköpings universitet

mattiasvillani.com  @matvil  mattiasvillani

# 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

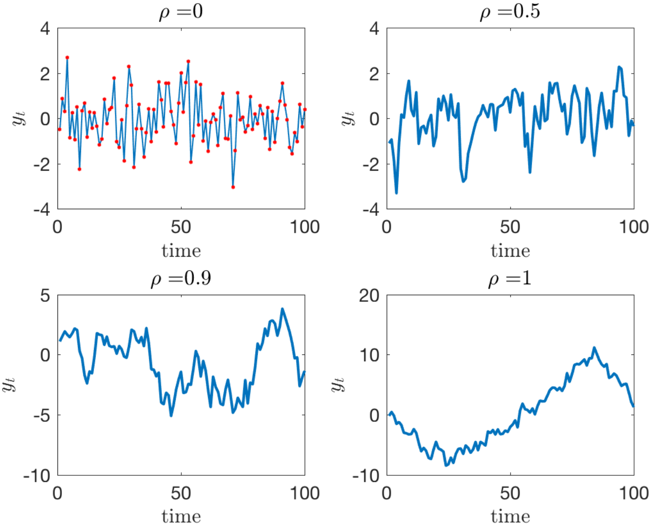
$$y_t|y_{t-1} \sim N\left(\rho y_{t-1}, \sigma^2\right).$$

- $AR(p)$ 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 $x_t \in \{1, 2\}$

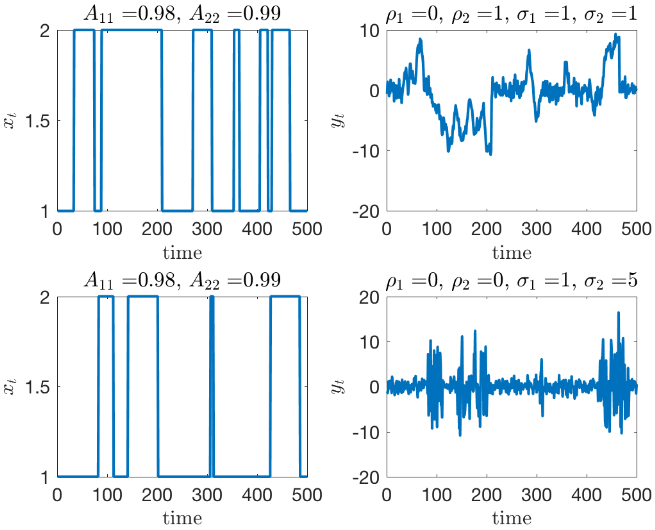
$$y_t = \begin{cases} \rho_1 y_{t-1} + \varepsilon_t, & \varepsilon_t \overset{iid}{\sim} N(0, \sigma_1^2) \quad \text{if } z_t = 1 \\ \rho_2 y_{t-1} + \varepsilon_t, & \varepsilon_t \overset{iid}{\sim} N(0, \sigma_2^2) \quad \text{if } z_t = 2 \end{cases}$$

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

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

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

# Hidden Markov models



# Time varying parameter models

- Smoothly **time varying parameter model**

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

- The persistence parameter $\rho$ 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**

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

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

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

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

- And so on ...

# Unobserved components models

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

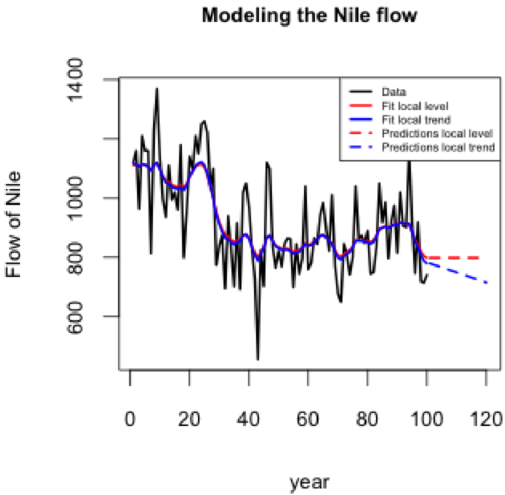
$$
\begin{aligned}
y_t &= \mu_t + \varepsilon_t \qquad \varepsilon_t \overset{iid}{\sim} N\left(0, \sigma_\varepsilon^2\right) \\
\mu_t &= \mu_{t-1} + \nu_t \qquad \nu_t \overset{iid}{\sim} N\left(0, \sigma_\nu^2\right)
\end{aligned}
$$

- **Local trend model**

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

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

# Local trend model is a state space model

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

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

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

- **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

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

# The posterior distribution of the state

- The linear Gaussian state-space (LGSS) model

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

$$\text{State eq:} \quad x_t = Ax_{t-1} + Bu_t + \nu_t \qquad \nu_t \overset{iid}{\sim} N(0, \Omega_\nu)$$

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

$$p(x_t | y_1, ..., y_T, u_1, ..., u_T)$$

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

$$p(x_1, ..., x_T | y_1, ..., y_T, u_1, ..., u_T)$$

- More on this later.

# Model structure

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

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

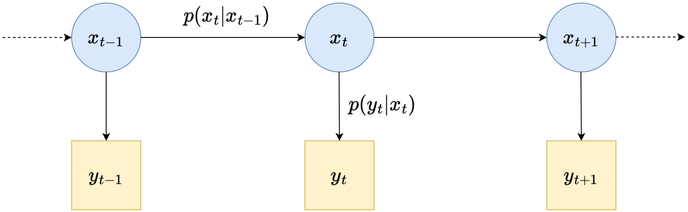
$$\text{State eq:} \quad x_t = Ax_{t-1} + Bu_t + \nu_t \qquad \nu_t \overset{iid}{\sim} N(0, \Omega_\nu)$$

- Note 1: $x_t$ is first order Markov:

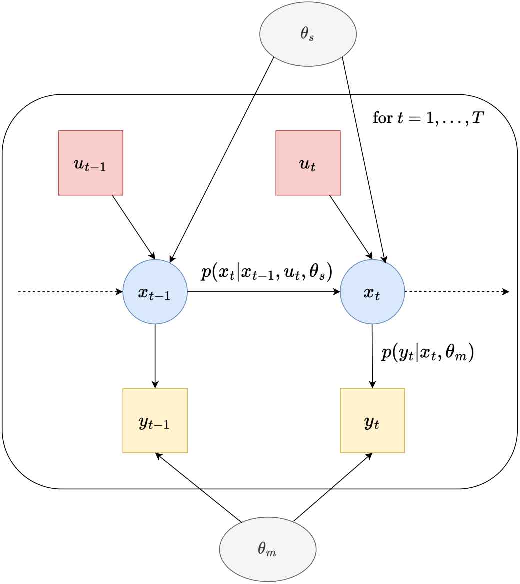
$$p(x_t | x_{t-1}, ..., x_1) = p(x_t | x_{t-1})$$

- Note 2: **Conditional** on $x_t$, $y_t$ is independent of past observations and states.

# State-space model as graphical model



# State-space model as graphical model



# The filtering distribution

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

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

- **Filtering = Posterior** for $\mathrm{x}_t$, **after** observing $\boldsymbol{y}_t$

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

- **Prior** for $\mathrm{x}_t$, **before** observing $\boldsymbol{y}_t$

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

- Recall: **smoothing distribution** conditions on all data

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

# The Bayes filter

- We are now at time $t$.
- We have just given the control command $\mathrm{u}_t$.
- We have not yet observed $\mathrm{y}_t$.

- **Prior** beliefs at this stage:

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

- Now comes the observation $\mathrm{y}_t$.
- **Update** to **posterior** using **Bayes' theorem**:

$$\pi(\mathrm{x}_t) \propto p(\mathrm{y}_t | \mathrm{x}_t) \bar{\pi}(\mathrm{x}_t).$$

# The Bayes filter

- **Prediction step** (control update)

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

- **Measurement update step**

$$\pi(\mathrm{x}_t) \propto p(\mathrm{y}_t|\mathrm{x}_t)\bar{\pi}(\mathrm{x}_t).$$

# 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**

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

- **Measurement update**

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

- 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**

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

$$\text{State eq:} \quad x_t = Ax_{t-1} + Bu_t + \nu_t \qquad \nu_t \overset{iid}{\sim} N(0, \Omega_\nu)$$

- `Algorithm` **KalmanFilter**$(\mu_{t-1}, \Sigma_{t-1}, u_t, y_t)$
  - Prediction update: $\begin{cases} \bar{\mu}_t = A\mu_{t-1} + Bu_t \\ \bar{\Sigma}_t = A\Sigma_{t-1}A^T + \Omega_\nu \end{cases}$
  - Measurement update : $\begin{cases} K_t = \bar{\Sigma}_t C^T \left(C\bar{\Sigma}_t C^T + \Omega_\varepsilon\right)^{-1} \\ \mu_t = \bar{\mu}_t + K_t(y_t - C\bar{\mu}_t) \\ \Sigma_t = (I - K_t C)\bar{\Sigma}_t \end{cases}$
  - Return $\mu_t, \Sigma_t$

# The Kalman filter code - single update

```
function kalmanfilter_update(μ, Σ, u, y, A, B, C, Q, R)

    # Prediction step - moving state forward without new measurement
    μ̄ = A*μ + B*u;
    Σ̄ = A*Σ*A' + R;

    # Measurement update - updating the N(μ̄, Σ̄) prior with the new data point
    K = Σ̄*C' / (C*Σ̄*C' + Q); # Kalman Gain
    μ = μ̄ + K*(y - C*μ̄);
    Σ = ( I(length(μ)) - K*C )*Σ̄;

    return μ, Σ
end
```

# The Kalman filter code

```
function kalmanfilter(U, Y, A, B, C, Q, R, μₒ, Σₒ)

    # Prelims
    T = size(Y,1)
    n = length(μₒ)

    μ_all = zeros(T,n)
    Σ_all = zeros(n,n,T)

    # The Kalman iterations
    μ = μₒ
    Σ = Σₒ
    for t = 1:T
        μ, Σ = kalmanfilter_update(μ, Σ, U[t,:]', Y[t,:]', A, B, C, Q, R)
        μ_all[t,:] = μ
        Σ_all[:,:,t] = Σ
    end

    return μ_all, Σ_all
end
```

# 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

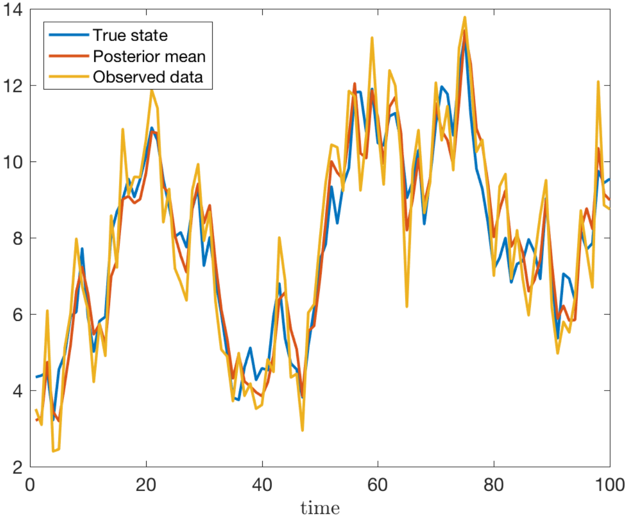
- The linear Gaussian state-space (LGSS) model

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

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

- 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