Statistical Data Analysis

Dr. Jana de Wiljes

14. November 2021

Universität Potsdam

Setting: parameter θ is a random variable and distributed according a known distribution

Bayes Theorem:

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(B|A) \cdot \mathbb{P}(A)}{\mathbb{P}(B)} \tag{1}$$

Goal: determine $\mathbb{P}(A|B)$

Example

Example: • consider a large car insurance that is categorising their clients in two types : Type 1 and Type (age, gender ...)

⇝ depending on the type one is categorised one payes more or less

- Probability for type 1 to report an $\underbrace{\text{accidents or defect}}$ is $\theta_1 = 0.4$
- Probability for type 2 to report 〃 is $\theta_2 = 0.1$

Let us now consider a person that could not directly be categorised

Which of the two car types fits best?

- Parameter space : $\Theta = \{\theta_1, \theta_2\}$
- S is random variable associated with number of accidents occurring in n years

$S \sim Bin(n, p)$

Data on person: $S = 2$ in $n = 10$ years

$$L(s;\theta_1) = \mathbb{P}_{\theta_1}[S = s] = \binom{n}{s}\theta_1^s(1-\theta_1)^{n-s} = \binom{10}{2}\cdot \underbrace{0.4}_{\theta_1}{}^2 \cdot 0.6^8 = 0.1209$$

Example

$$L(s, \theta_2) = \mathbb{P}_{\theta_2}[S=s] = \binom{n}{s} \theta_2^s (1-\theta_2)^{n-s} = \binom{16}{2} 0.1^2 \cdot 0.98 = 0.1937$$

$$L(s, \theta_2) > L(s, \theta_1)$$

$\rightsquigarrow$ ML estimator would yield to put this person in type 2.

Let us assume we have some valuable information, i.e.,

$$\boxed{q(\theta_1) = \mathbb{P}[\Theta = \theta_1] = 0.9} \qquad q(\theta_2) = \mathbb{P}[\Theta = \theta_2] = 0.1$$

$\rightsquigarrow$ in other words Type 1 is much more likely

a priori Distribution

$$\mathbb{P}[S=s] = \boxed{\mathbb{P}[\Theta = \theta_1]} \cdot \boxed{\mathbb{P}[S=s \mid \Theta = \theta_1]} + \mathbb{P}[\Theta = \theta_2] \cdot \mathbb{P}[S=s \mid \Theta = \theta_2]$$

$$= q(\theta_1) \binom{n}{s} \theta_1^s (1-\theta_1)^{n-s} + q(\theta_2) \binom{n}{s} \theta_2^2 (1-\theta_2)^{n-s}$$

Example

For $s=2$ we have

$$\mathbb{P}[S=2] = 0.9 \cdot 0.1209 + 0.1 \cdot 0.1937 = 0.1282$$

We are interested in the so called a posteriori distribution:

$$q(\theta_1 | s) = \boxed{\mathbb{P}[\theta = \theta_1 | S = s] = \frac{\mathbb{P}[\theta = \theta_1 \cap S = s]}{\mathbb{P}[S=s]} = \frac{\overset{\text{prior}}{\mathbb{P}[\theta = \theta_1]}\, \mathbb{P}[S = s | \theta = \theta_1]}{\mathbb{P}[S=s]}}$$

$$q(\theta_1 | 2) = \frac{0.9 \cdot 0.1209}{0.1282} = 0.8486$$

$$q(\theta_2 | 2) = 1 - q(\theta_1 | 2) = 0.1513$$

$$q(\theta_1 | 2) > q(\theta_2 | 2)$$

Def: The a-posteriori-distribution of θ is the conditional distribution given the information $X_1 = x_1, \dots, X_n = x_n$, i.e.,

$$q(\theta_i | x_1, \dots, x_n) := \mathbb{P}[\theta = \theta_i | \underbrace{X_1 = x_1, \dots, X_n = x_n}_{\text{Data}}], \quad i = 1, 2, \dots$$

$$\overset{\text{iid}}{=} \frac{\mathbb{P}[\Theta = \Theta_i] \cdot \mathbb{P}[X_1 = x_1, \dots, X_n = x \mid \Theta = \Theta_i]}{\mathbb{P}[X_1 = x_1, \dots, X_n = x_n]}$$

$$\overset{\text{iid } X_i}{=} \frac{q(\Theta_i)\, h_{\Theta_i}(x_1) \cdot \ldots \cdot h_{\Theta_i}(x_n)}{\sum_j q(\Theta_j)\, h_j(x_1) \cdot \ldots \cdot h_{\Theta_j}(x_n)}$$

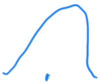
$$h_{\Theta_i}(x_n) = \mathbb{P}(X_n = x_n | \theta_i)$$

A-posteriori-distribution

Bayes method

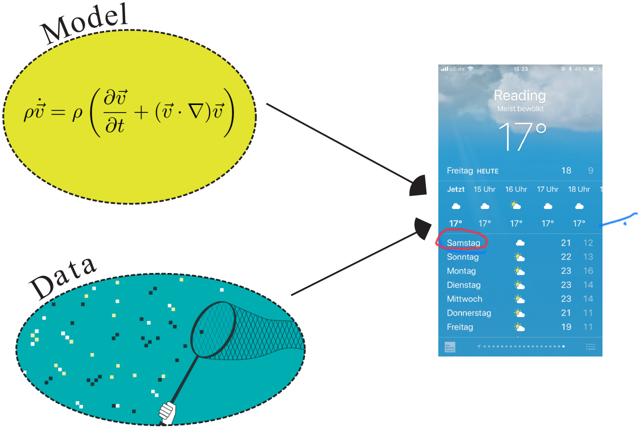
Def: The Bayes estimator is defined as the expectation of the a-posteriori-distribution

$$\widehat{\theta}_{\text{Bayes}} = \sum_i \theta_i \underbrace{q(\theta_i|x_1,\ldots,x_n)}_{\text{posterior}}$$



What is Data Assimilation?

Data Assimilation

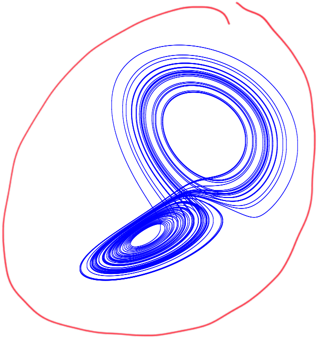


Uncertainty?

A toy atmospheric model

Lorenz equations:

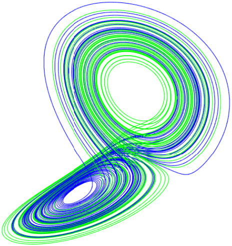
$$\dot{x} = \sigma(y - x)$$
$$\dot{y} = x(\rho - z) - y$$
$$\dot{z} = xy - \beta z$$



Uncertainty in initial conditions

Lorenz equations:

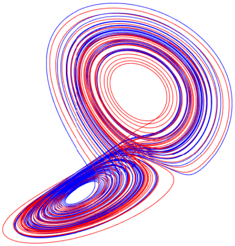
$$\dot{x} = \sigma(y - x)$$
$$\dot{y} = x(\rho - z) - y$$
$$\dot{z} = xy - \beta z$$



Uncertainty in parameters

Lorenz equations:

$$\dot{x} = \sigma(y - x)$$
$$\dot{y} = x(\rho - z) - y$$
$$\dot{z} = xy - \beta z$$



Lorenz equations:

n is time

$$
\begin{aligned}
x_n &= x_{n-1} + [\sigma(y_{n-1} - x_{n-1})]\mathrm{dt} \\
y_n &= y_{n-1} + [x_{n-1}(\rho - z_{n-1}) - x_{n-1}]\mathrm{dt} \\
z_n &= z_{n-1} + [x_{n-1}y_{n-1} - \beta z_{n-1}]\mathrm{dt}
\end{aligned}
$$

The Model

Model (deterministic)

Evolution equation

$$z_n = \Psi(z_{n-1}, \lambda) \quad + \xi_n$$

where

$$z_0 \sim \mathcal{N}(m_0, C_0)$$

$$\xi_n \sim N(0, B)$$

Model

Evolution equation

$$z_n = \Psi(z_{n-1}, \lambda) + \xi_{n-1}$$

where

$$z_0 \sim \mathcal{N}(m_0, C_0)$$

$$\xi_n \sim \mathcal{N}(0, B) \quad \text{i.i.d.} \quad \forall\ n$$

Parameter estimation

Augmented state space

$$\begin{aligned} z_n &= \Psi(z_{n-1}, \lambda_{n-1}) + \xi_{n-1} \\ \lambda_n &= \lambda_{n-1} \end{aligned}$$

where

$$\begin{aligned} [z_0, \lambda_0]^\top &\sim \mathcal{N}(m_0, C_0) \\ \xi_n &\sim \mathcal{N}(0, B) \quad \text{i.i.d.} \quad \forall\ n \end{aligned}$$

Observations

Observations

Partial and noisy data:

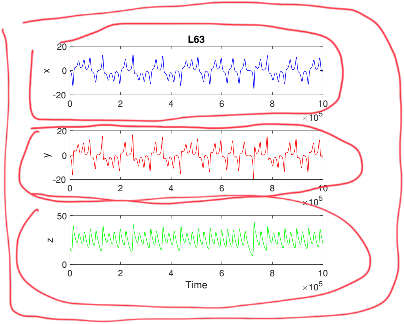
observation operator

$$y_n = h(z_n) + \eta_n$$

where

$$\eta_n \sim \mathcal{N}(0, R) \quad \text{i.i.d.} \quad \forall\, n$$

Example



Lorenz equations:

$$x_n = x_{n-1} + [\sigma(y_{n-1} - x_{n-1})]\mathrm{dt}$$
$$y_n = y_{n-1} + [x_{n-1}(\rho - z_{n-1}) - x_{n-1}]\mathrm{dt}$$
$$z_n = z_{n-1} + [x_{n-1}y_{n-1} - \beta z_{n-1}]\mathrm{dt}$$

The Math behind it...

Conditional probability

Definition (Conditional probability)

For a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and events $A, B \in \mathcal{F}$ the conditional probability of B given A is

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A, B)}{\mathbb{P}(B)}.$$

Bayes theorem

Theorem (Bayes)

For a given probability space $(\Omega, \mathcal{F}, \mathbb{P})$ *the following holds for two events* A *and* B *in* $\mathcal{F}$

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(B|A)\mathbb{P}(A)}{\mathbb{P}(B)}.$$

Bayesian data assimilation ansatz

Likelihood

$$\mathbb{P}(\text{Model}|\text{Obs}) = \frac{\mathbb{P}(\text{Obs}|\text{Model})\mathbb{P}(\text{Model})}{\mathbb{P}(\text{Obs})}$$

Bayesian data assimilation ansatz

$$\mathbb{P}(\text{Model}|\text{Obs}) \propto \mathbb{P}(\text{Obs}|\text{Model})\mathbb{P}(\text{Model})$$

Bayesian data assimilation for densities

y_{n+1}

$$
\begin{aligned}
\pi(z_{n+1}|y_{1:n+1}) &= \pi(z_{n+1}|y_{1:n}, y_{n+1}) \\
&= \frac{\pi(y_{n+1}|\cancel{y_{1:n}}, z_{n+1})\pi(z_{n+1}|y_{1:n})}{\pi(y_{n+1}|y_{1:n})} \\
&= \frac{\pi(y_{n+1}|z_{n+1})\pi(z_{n+1}|y_{1:n})}{\pi(y_{n+1}|y_{1:n})}
\end{aligned}
$$

- prior

likelihood

normalising

$$
\implies \pi(z_{n+1}|y_{1:n+1}) \propto \pi(y_{n+1}|z_{n+1})\pi(z_{n+1}|y_{1:n}) \qquad (2)
$$

Special case

Linear model: Ψ is linear, e.g.,

$N(0, B)$

$$\boxed{z_n = A z_{n-1} + \xi_{n-1}} \tag{3}$$

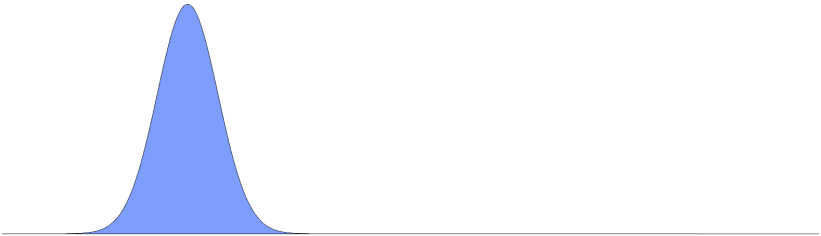
with $A \in \mathbb{R}^{N_z} \times \mathbb{R}^{N_z}$

$z_0 \sim \boxed{N(m_0, C_0)}$

Linear observation operator

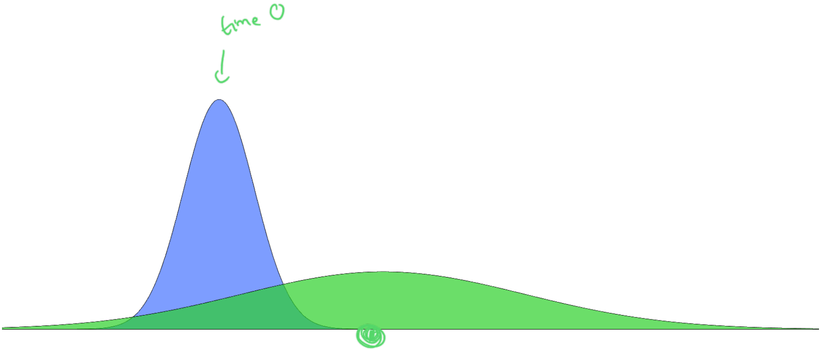
$$\boxed{h = H \quad \text{with} \quad H \in \mathbb{R}^{N_y} \times \mathbb{R}^{N_y}}$$

Linear model



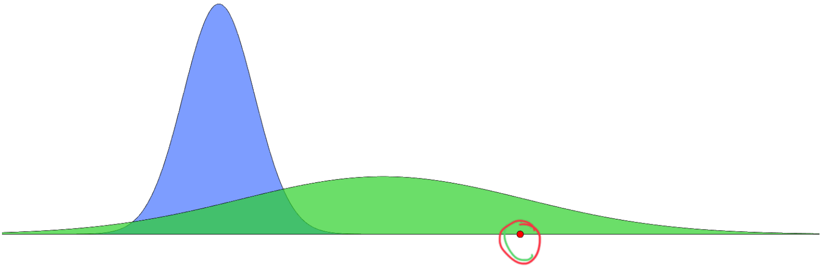
Initial distribution: $z_0 \sim \mathcal{N}(m_0, C_0)$

Linear model



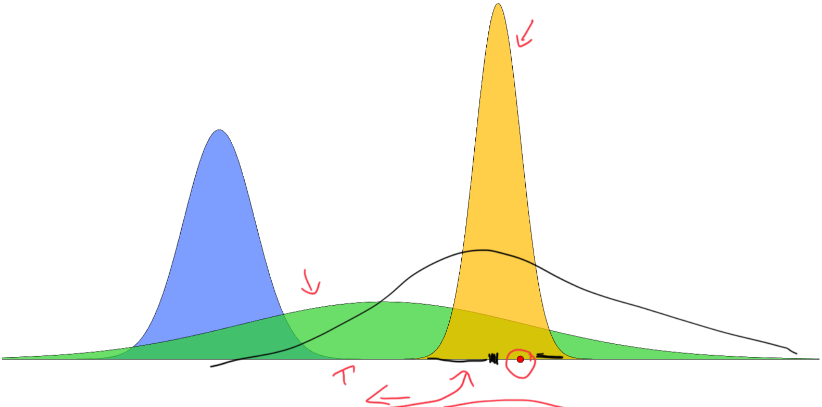
Prior distribution: $\mathcal{N}(\hat{m}_1, \hat{C}_1)$

Linear model



Likelihood: $\mathcal{N}(H\hat{m}_1, R)$

Linear model



Posterior: $\mathcal{N}(m_1^{a}, C_1^{a}) \propto \mathcal{N}(H\hat{m}_1, R)\,\mathcal{N}(\hat{m}_1^{f}, \hat{C}_1^{f})$ f

Algorithms

Kalman filter

Two steps:

Kalman filter

Two steps:

- Forecast: $(m_n, C_n) \mapsto (\hat{m}_{n+1}, \hat{C}_{n+1})$

Kalman filter

Two steps:

- Forecast: $(m_n, C_n) \mapsto (\hat{m}_{n+1}, \hat{C}_{n+1})$
- Analysis: $(\hat{m}_{n+1}, \hat{C}_{n+1}) \mapsto (m_{n+1}, C_{n+1})$

Kalman filter

Two steps:

- Forecast: $(m_n, C_n) \mapsto (\hat{m}_{n+1}, \hat{C}_{n+1})$
- Analysis: $(\hat{m}_{n+1}, \hat{C}_{n+1}) \mapsto (m_{n+1}, C_{n+1})$

Forecast formulas

$$\hat{m}_{n+1} = A m_n$$
$$\hat{C}_{n+1} = A C_n A^\top + B$$

Kalman filter

Two steps:

- Forecast: $(m_n, C_n) \mapsto (\hat{m}_{n+1}, \hat{C}_{n+1})$
- Analysis: $(\hat{m}_{n+1}, \hat{C}_{n+1}) \mapsto (m_{n+1}, C_{n+1})$

Forecast formulas

$$\hat{m}_{n+1} = A m_n$$
$$\hat{C}_{n+1} = A C_n A^\top + B$$

$\xi \sim N(0, B)$

Analysis formulas

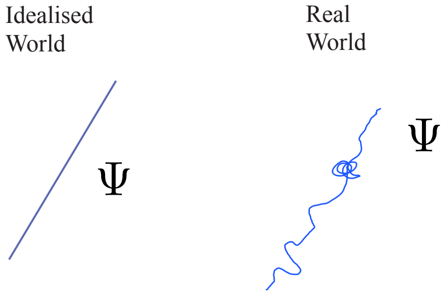
$$m_{n+1} = \hat{m}_{n+1} - K_{n+1}(H\hat{m}_{n+1} - y_{n+1})$$
$$C_{n+1} = \hat{C}_{n+1} - K_{n+1} H \hat{C}_{n+1}$$

Kalman gain

$$K_{n+1} = \hat{C}_{n+1} H^\top (R + H \hat{C}_{n+1} H^\top)^{-1}$$

Kalman filter

Reality check

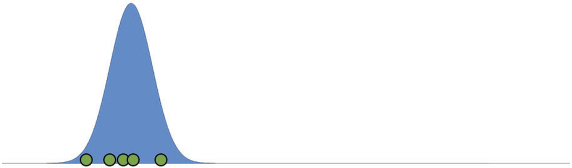


Problem: Kalman Filter is not applicable anymore

Approach: use approximative Algorithms, e.g.,

- Extended Kalman Filter: linearize model function

Monte Carlo approximation



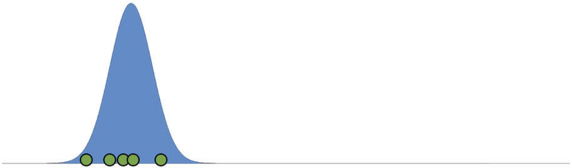
Approach: approximative via empirical measure

$$\pi(z_n|y_{1:n}) = \frac{1}{M}\sum_{i=1}^{M}\delta(z - z_n^i)$$

where

$$z_n^i \sim \pi(z_n|y_n)$$

Ensemble Kalman filter

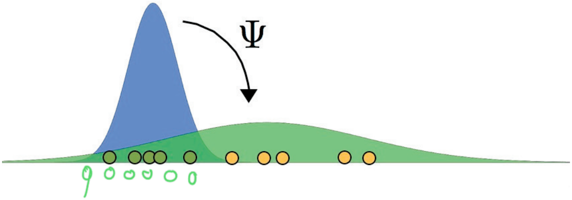


$$\mathcal{N}(m_0, C_0) \text{ with } m_0 \approx \frac{1}{M}\sum_{i=1}^{M} z_0^i$$

$$C_0 \approx \frac{1}{M}\sum_{i=1}^{M}(z_0^i - m_0)(z_0^i - m_0)^\top$$

M-1

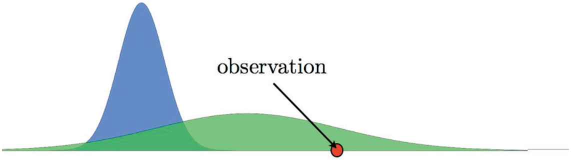
Ensemble Kalman filter



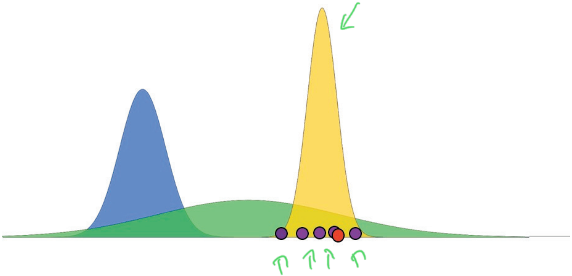
$$\mathcal{N}(\hat{m}_1, \hat{C}_1) \text{ with } \hat{m}_1 \approx \frac{1}{M}\sum_{i=1}^{M} \hat{z}_1^i = \frac{1}{M}\sum_{i=1}^{M} \Psi(z_0^i)$$

$$\hat{C}_1 \approx \frac{1}{M}\sum_{i=1}^{M} (\hat{z}_1^i - \hat{m}_1)(\hat{z}_1^i - \hat{m}_1)^\top$$

Ensemble Kalman filter



Ensemble Kalman filter



$$\mathcal{N}(m_1, C_1) \text{ with } m_1 \approx \frac{1}{M}\sum_{i=1}^{M} z_1^i$$

$$C_1 \approx \frac{1}{M}\sum_{i=1}^{M}(z_1^i - m_1)(z_1^i - m_1)^\top$$