Statistical Data Analysis

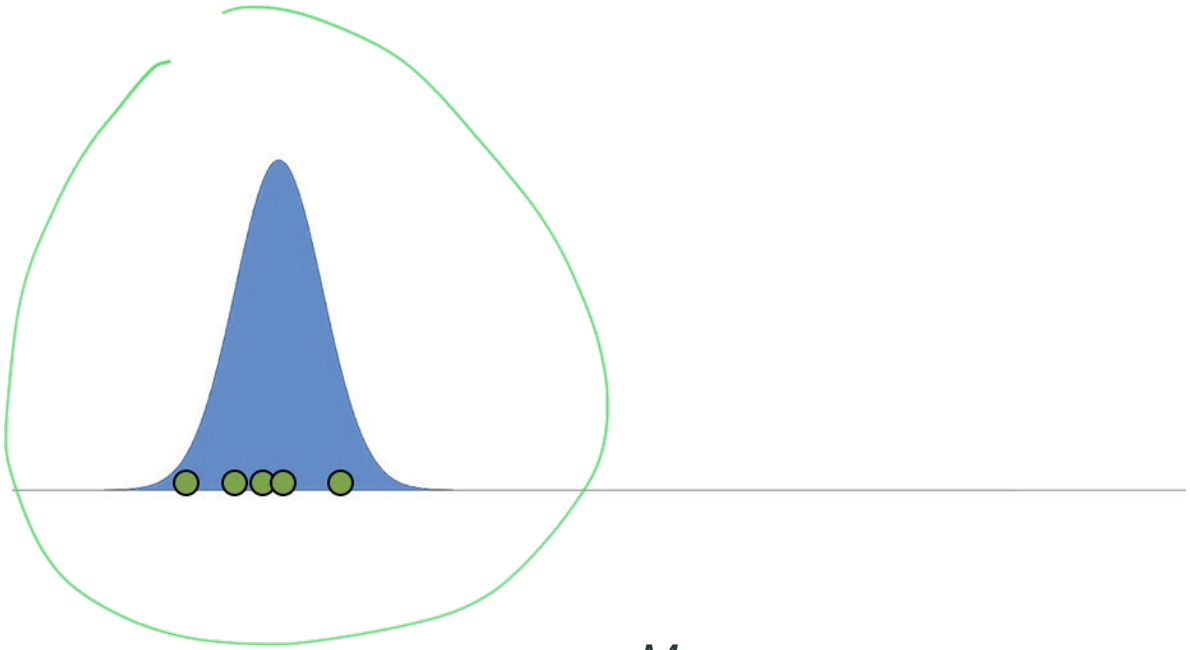
Dr. Jana de Wiljes

16. November 2021

Universität Potsdam

Lecture 12:30
next week
on tuesday

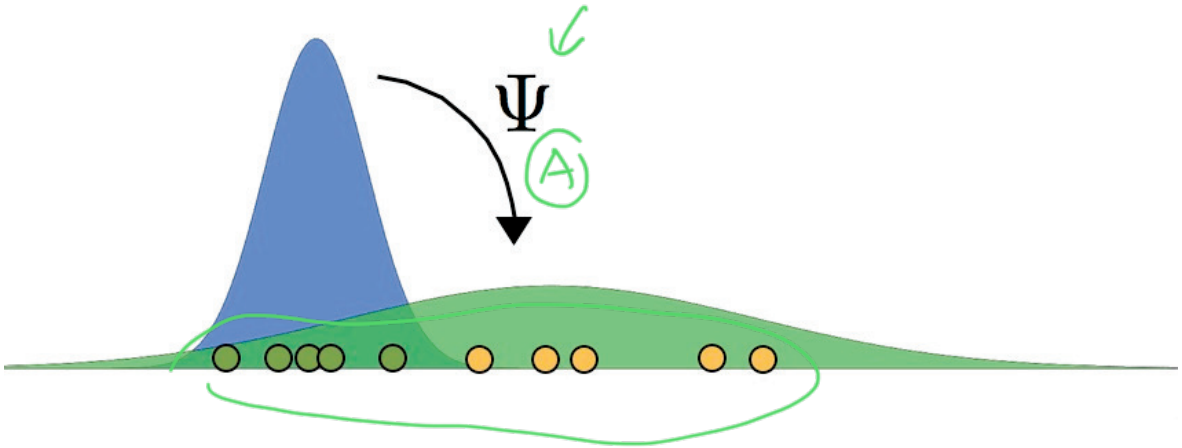
Ensemble Kalman filter



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

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

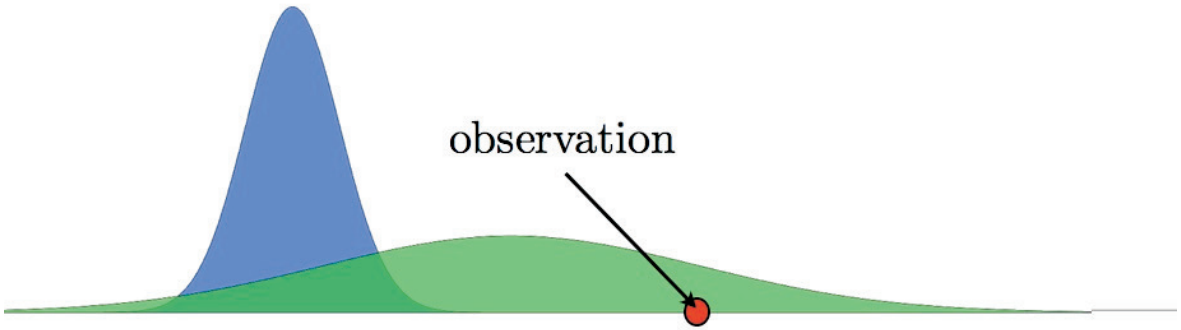
Ensemble Kalman filter



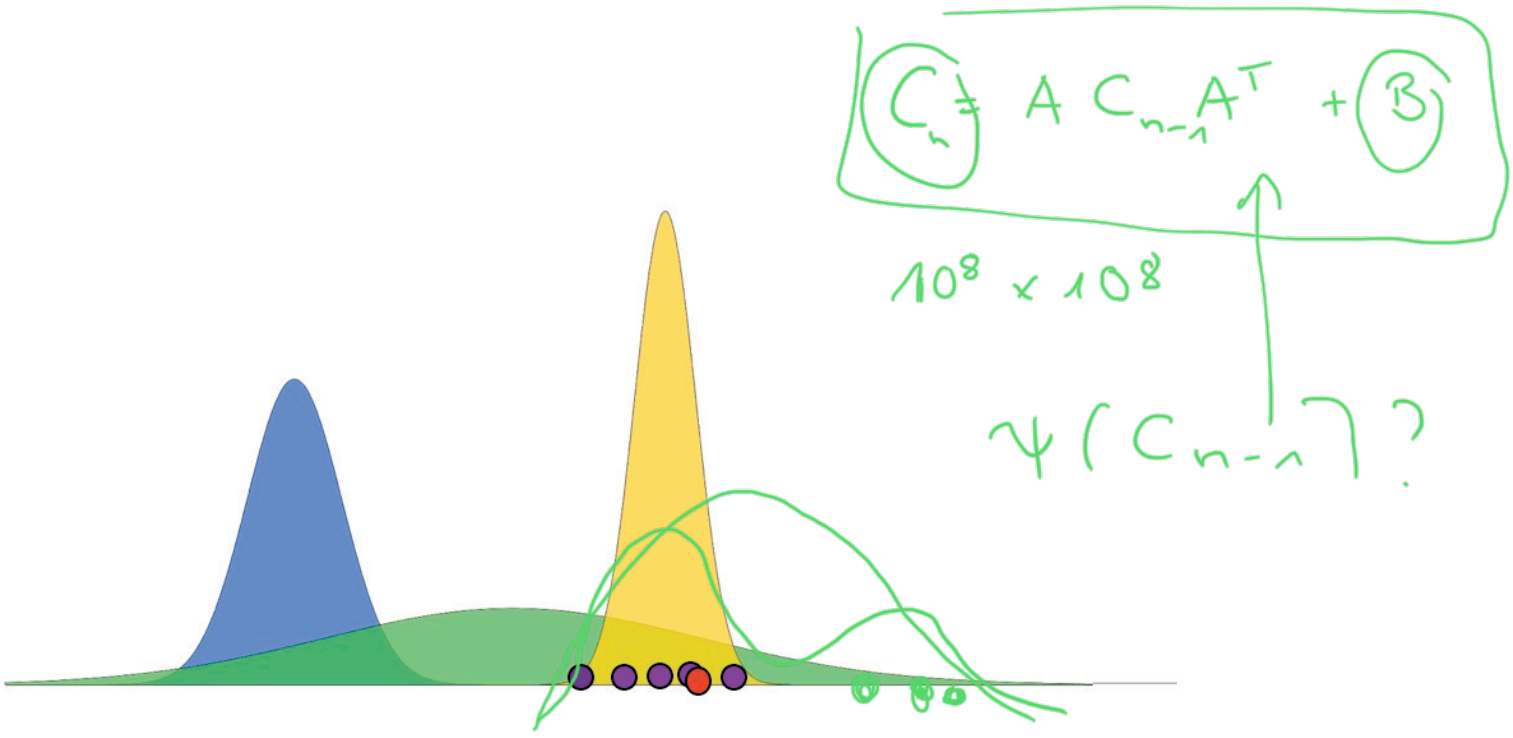
$$\mathcal{N}(\hat{m}_1, \hat{C}_1) \text{ with } \hat{m}_1 \approx \frac{1}{M}\sum_{i=1}^{M} \hat{z}_1^i = \frac{1}{M}\sum_{i=1}^{M} \Psi(z_0^i)$$

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

Ensemble Kalman filter



Ensemble Kalman filter



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

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

Ensemble Kalman filter

Goal: approximate $\pi(z_n|y_{1:n})$

Ensemble Kalman filter

Goal: approximate $\pi(z_n|y_{1:n})$

Approach: propagate samples $\hat{z}^i_{n+1}$ with Kalman formula

$$z^i_{n+1} = \hat{z}^i_{n+1} - K_{n+1}(H\hat{z}^i_{n+1} - \tilde{y}^i_{n+1})$$

Ensemble Kalman filter

Goal: approximate $\pi(z_n | y_{1:n})$

Approach: propagate samples $\hat{z}^i_{n+1}$ with Kalman formula

$$z^i_{n+1} = \hat{z}^i_{n+1} - K_{n+1}(H\hat{z}^i_{n+1} - \tilde{y}^i_{n+1})$$

Need: perturbed observations

$$\tilde{y}^i_{n+1} = y_{n+1} + \epsilon^i_{n+1}$$

with $\epsilon^i_{n+1} \sim \mathcal{N}(0, R)$ i.i.d. to get the correct mean and covariance in the linear case for $M \to \infty$

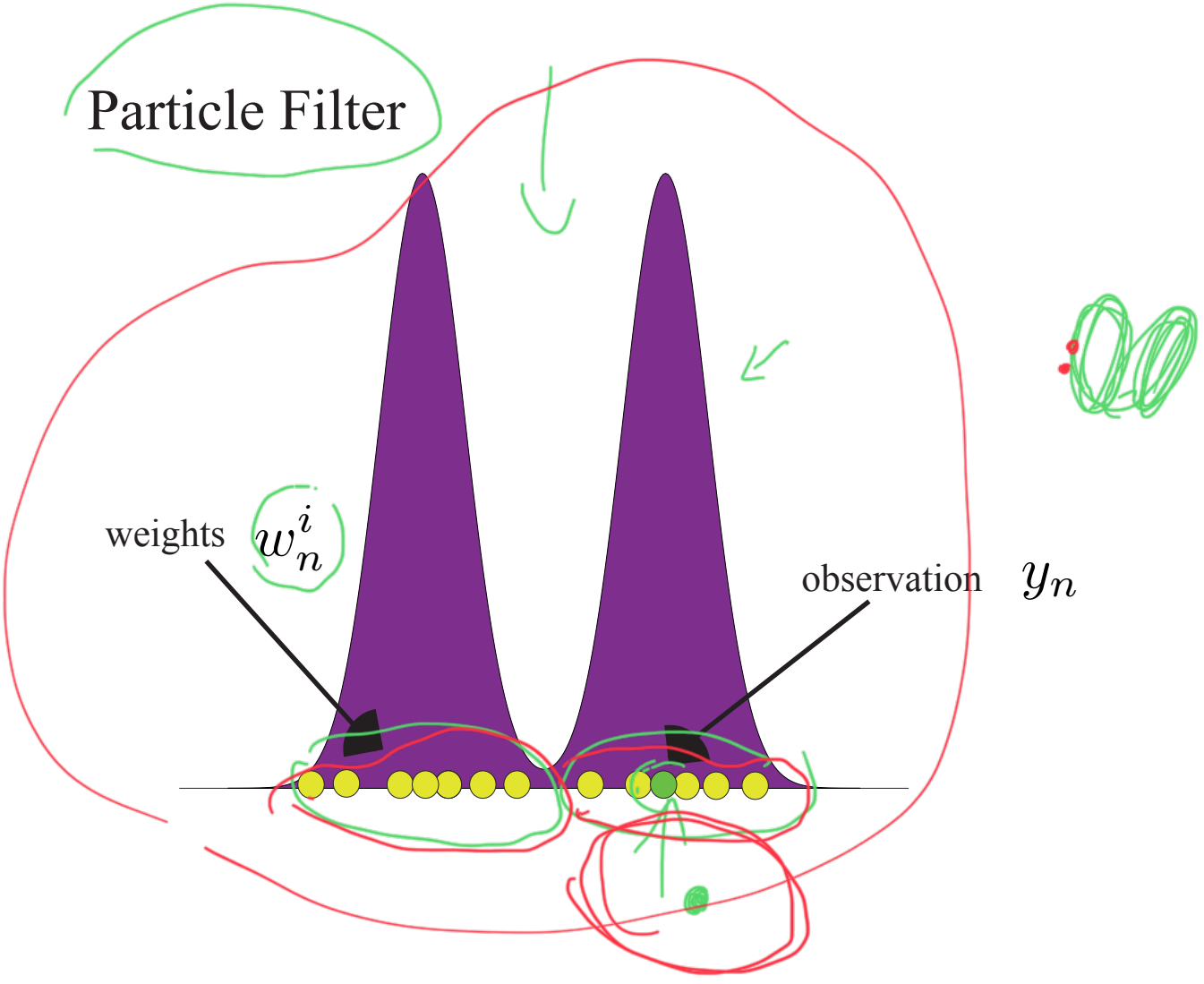
Ensemble Kalman filter

Works well in practice: e.g., EnKF is used for operational NWP for z_n^i with dimension 10^8 only using $M = 100$

Yet: mathematical foundation largely missing

Recent study: accuracy results for EnKF for idealized setting: $H = Id$ and observational error small

Particle filter



Particle filter

Problem: sampling from $\pi(z_n|y_{1:n})$ to approximate posterior via

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

Idea: sampling from $\pi(z_n|y_{1:n-1})$ instead i.e.,

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

Bayes:

$$\pi(z_{n+1}|y_{1:n}) \propto \pi(y_n|z_n)\pi(z_n|y_{1:n-1}) \tag{1}$$

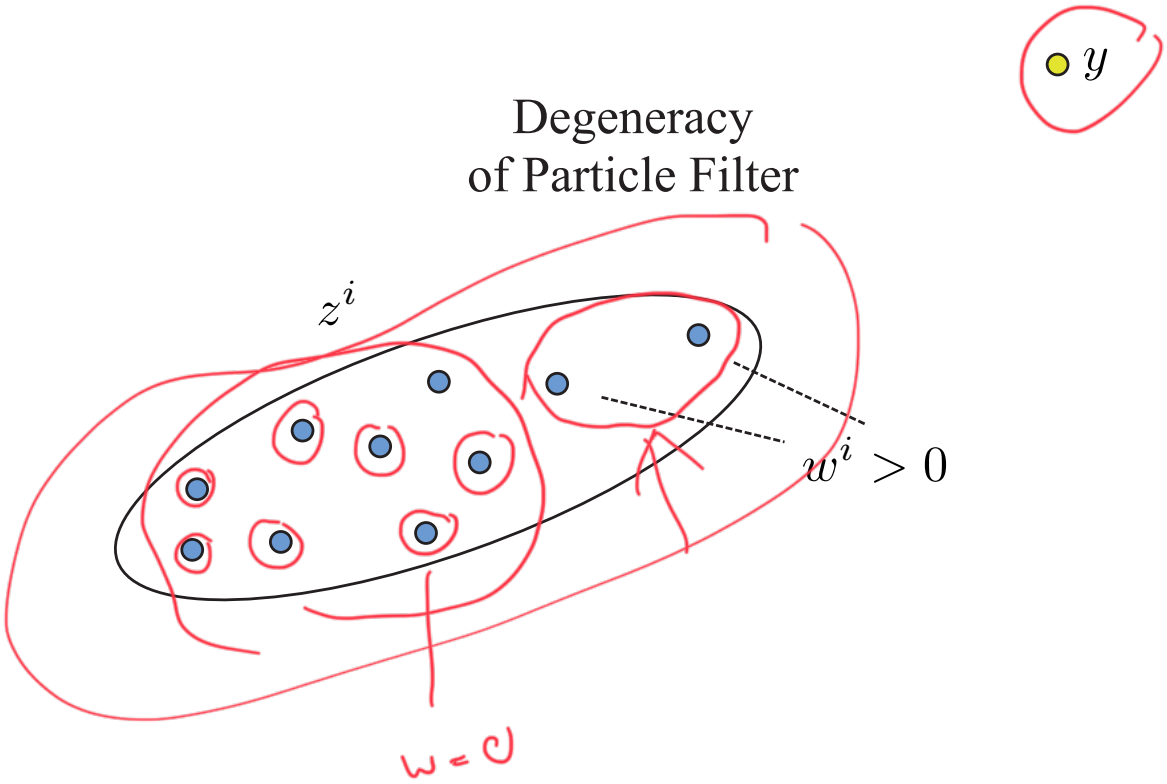
Weighting: unnormalized weights

$$\tilde{w}_n^i = \pi(y_n|z_n^i)w_{n-1}^i \text{ with } w_0^i = \frac{1}{M}$$

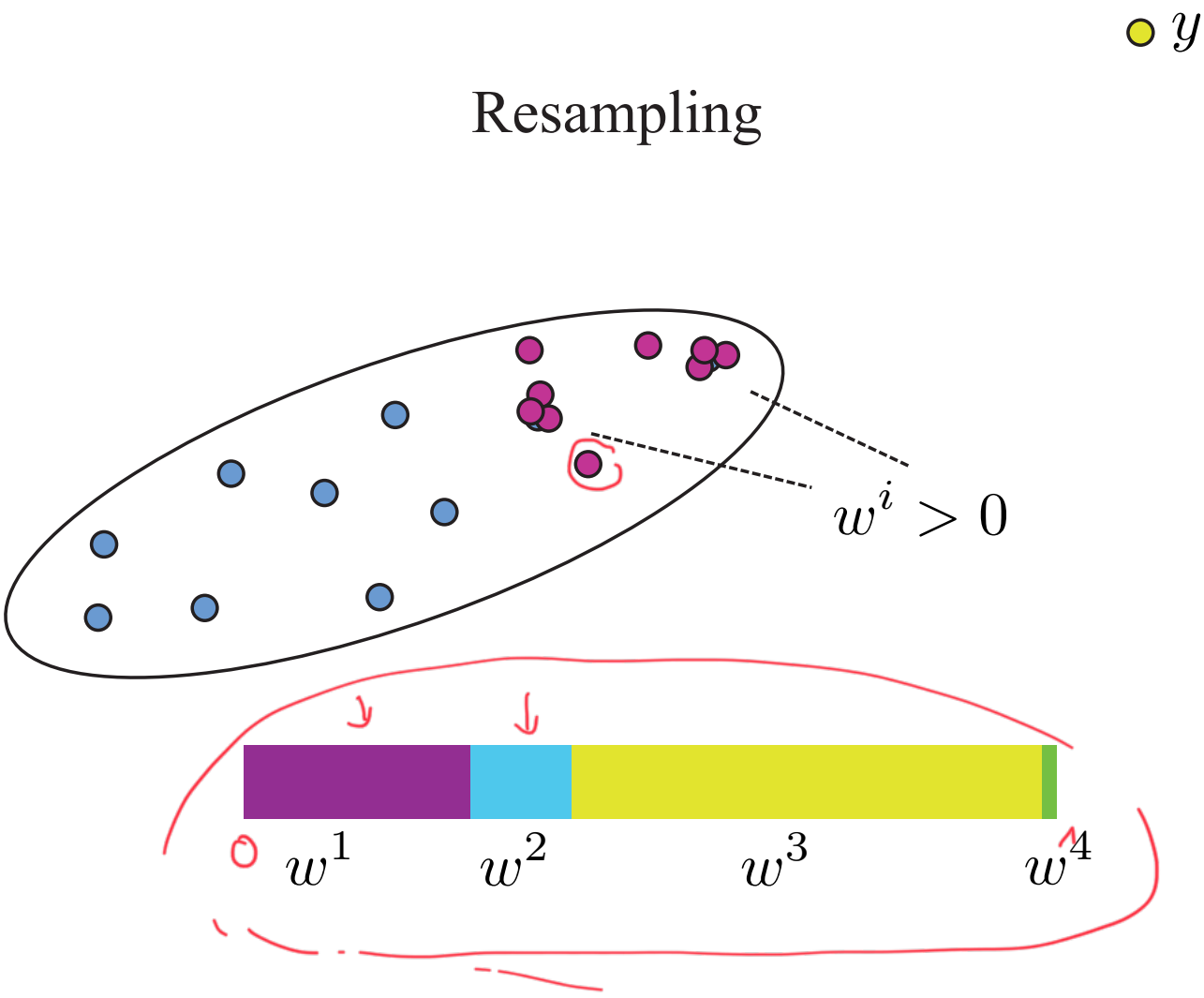
and normalized weights

$$w_n^i = \frac{\tilde{w}_n^i}{\sum_{j=1}^{M}\tilde{w}_n^j}$$

Particle collapse



Resampling



Resampling

Problem: weights w_n^i become very small

Ansatz: resampling

Input: w_n^i

For$(k = 1 : M)$

1. Draw a number $u \in [0, 1]$ from the uniform distribution $U[0, 1]$
2. Compute $i^* \in \{1, \ldots, M\}$ which satisfies

$$i^* = \arg\min_{i \geq 1} \sum_{j=1}^{i} w_j \geq u \tag{2}$$

3. Set $\xi_{i^*} = \xi_{i^*} + 1$

Return ξ_i

Still a lot of challenges....

Ansatz: approximative via empirical measure

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

where

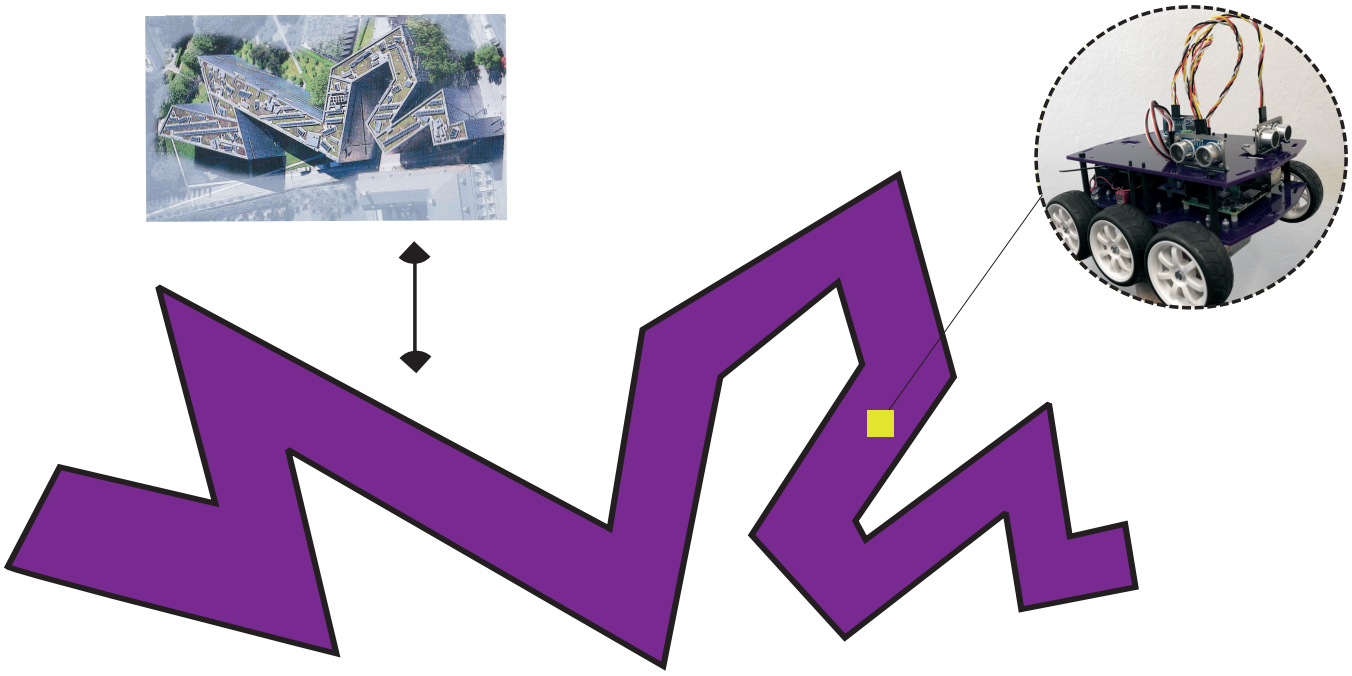
$$z_n^i \sim \pi(z_n|y_n)$$

Monte Carlo approximation leads to a variety of filters e.g.,

- Particle filters (curse of dimensionality)
- Ensemble Kalman filter (underlying Gaussian assumption)

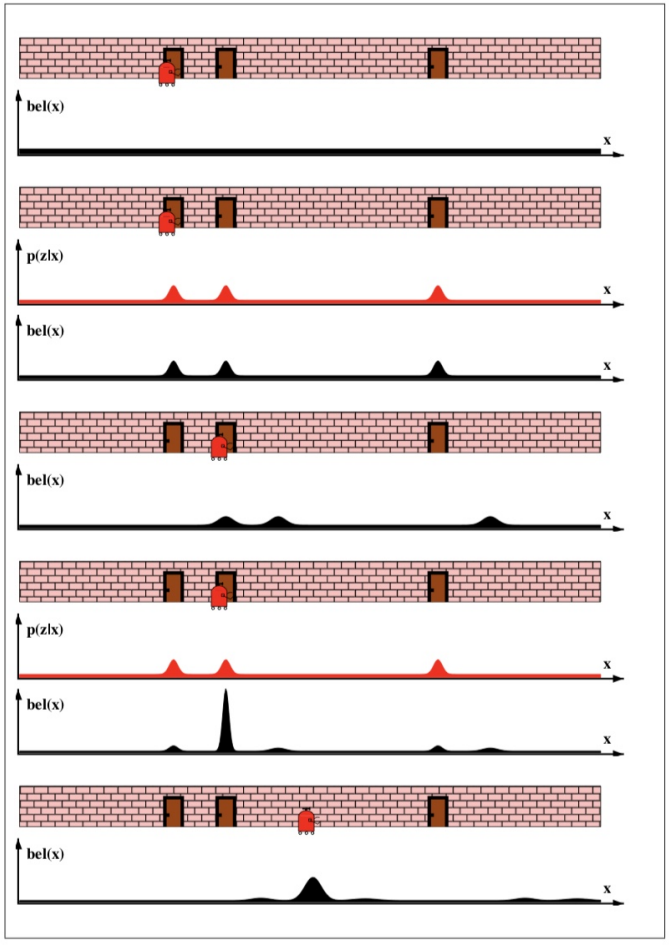
Same Math different Applications

Robotics

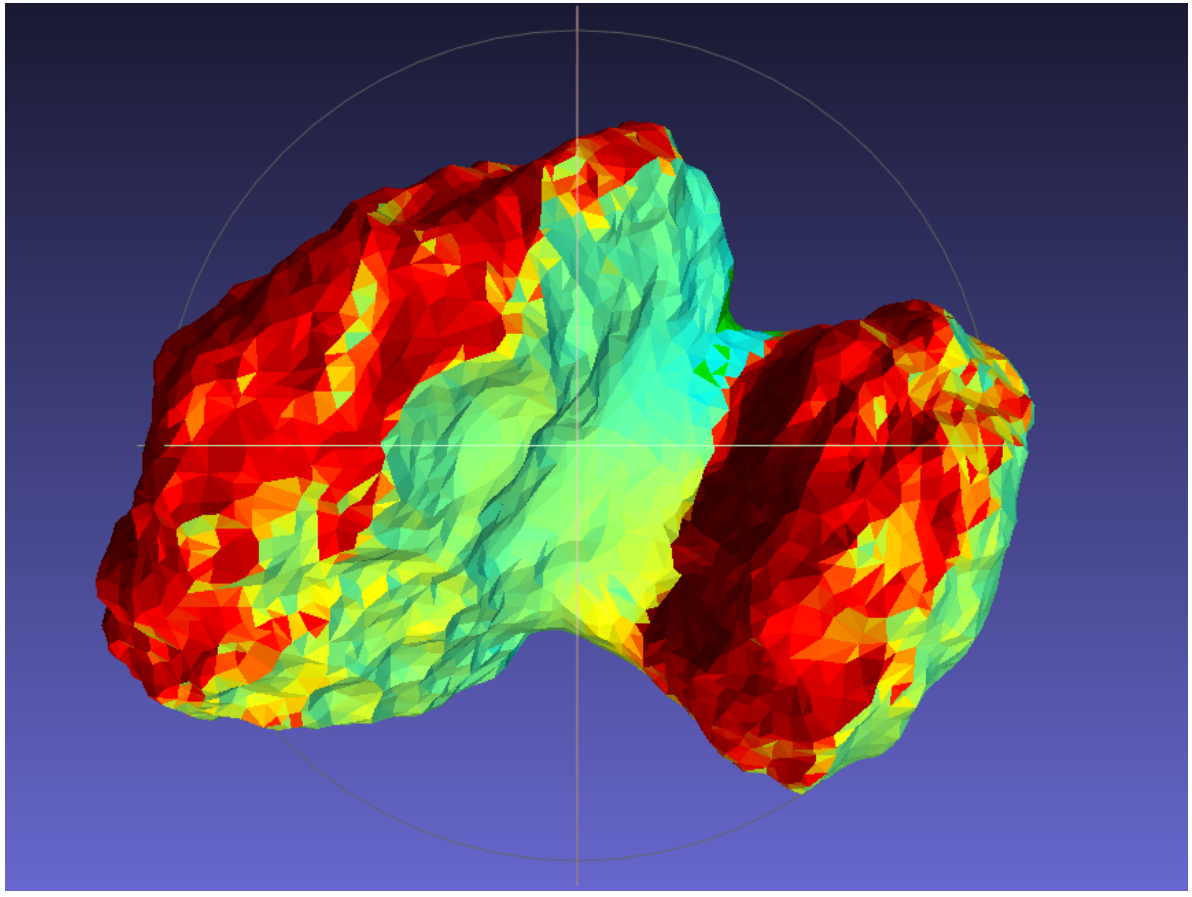


Simultaneous state and parameter estimation

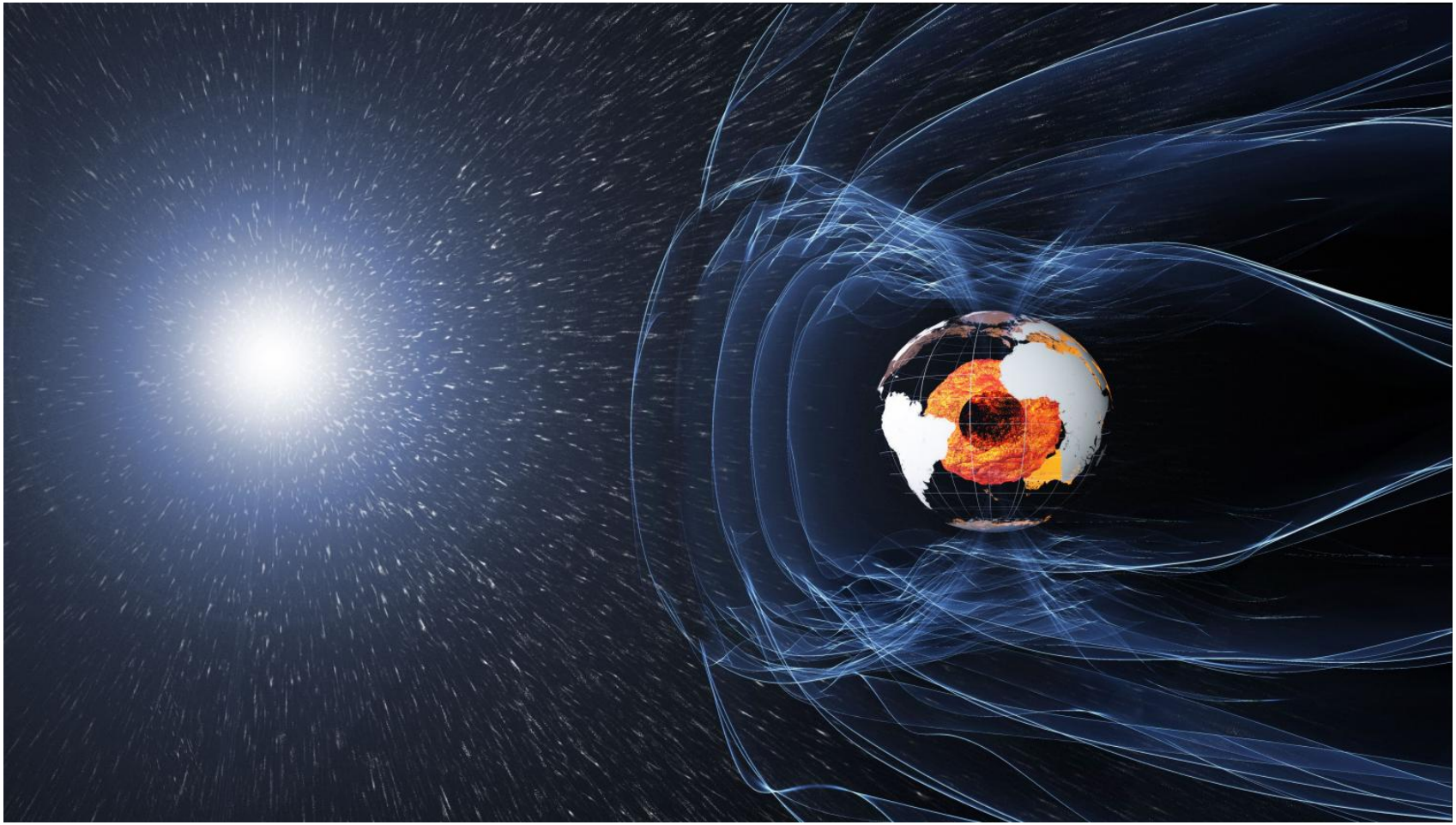
Robotics



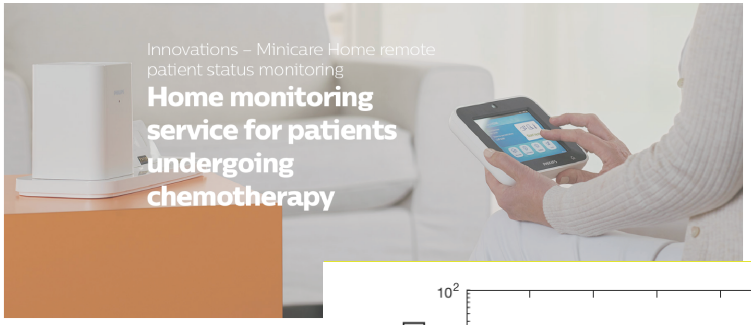
Thermophysical modeling

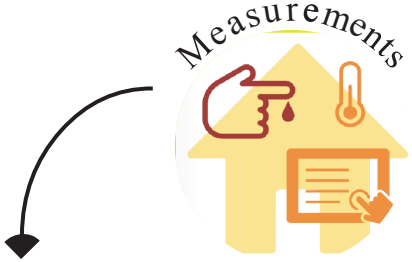


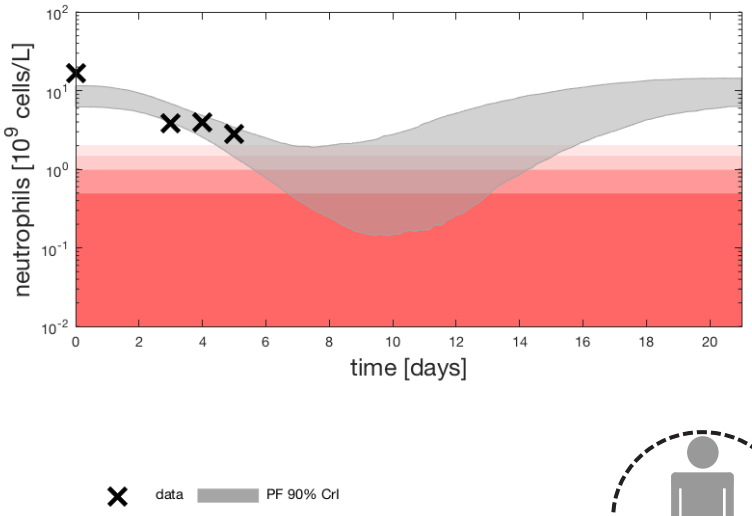
Space weather

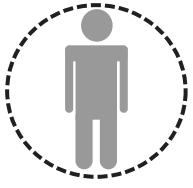


Pharmacology







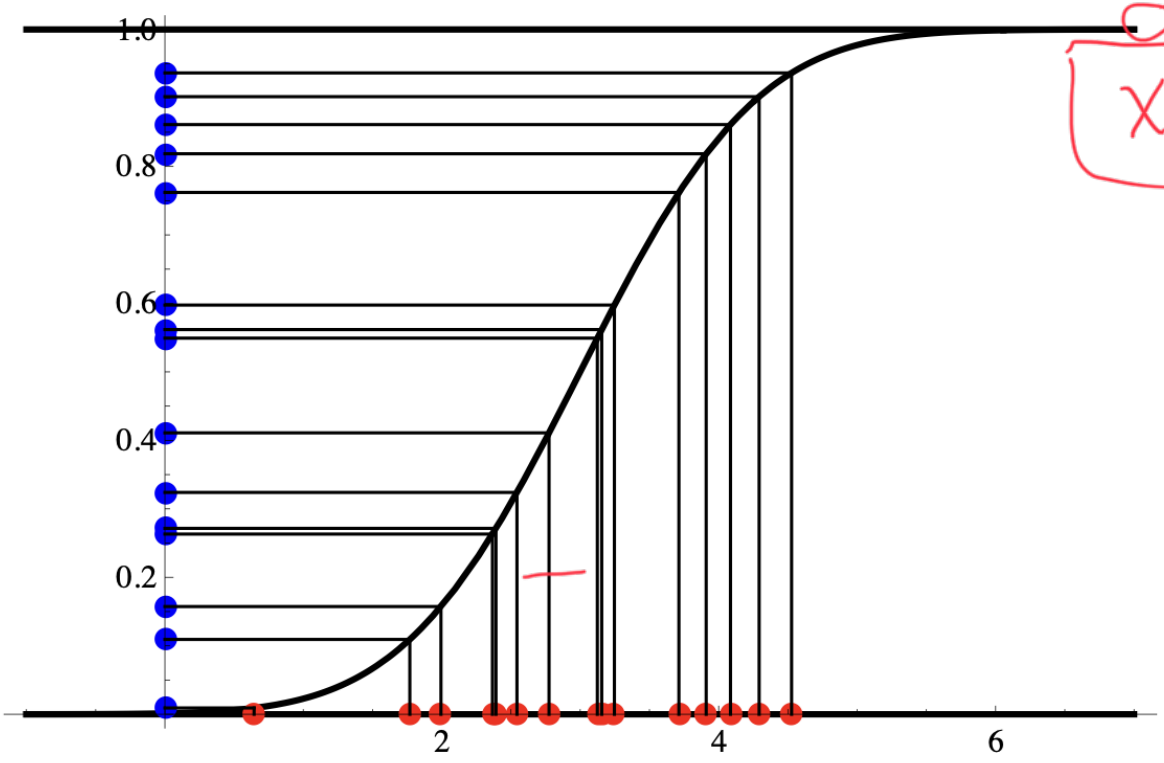


Personalised medicine

Maximum-spacing method

Maximum-spacing method

Lemma: Let the cdf F_θ be continuous and strictly monoton increasing. Under $\mathbb{P}_\theta$ the random variables $F_\theta(X_1), \ldots, F_\theta(X_n)$ are independent and uniformly distributed on the $(0, 1)$ interval.



Proof

Proof: The indepence follows from the independence of $X_1, \dots, X_n$

to show: $F_\theta(X_i)$ is uniformly distributed on $(0,1)$ $\forall i$

Note: $F_\theta(X_i)$ is assuming values in $(0,1)$

Let $x \in (0,1)$ then the following holds:

$$\mathbb{P}_\theta[F_\theta(X_i) \leq x] = \mathbb{P}_\theta[X_i \leq F_\theta^{-1}(x)] = F_\theta(F_\theta^{-1}(x)) = x$$

$\mathbb{R}$ $\Rightarrow F_\theta(X_i)$ uniform $\square$

$F_\theta(X_i): \Omega \to E = (0,1)$

Proof

Maximum-spacing method

Lemma: Let $z_1, \ldots, z_k \in [0,1]$ be number that are subject to the condition $z_1 + \cdots + z_k = 1$. Then

$$z_1 \cdot \cdots \cdot z_k \leq \frac{1}{k^k}. \tag{3}$$

Equality is attained only if all the numbers all the numbers are equal to $\frac{1}{k}$.

Beispiel $k = 2$

$z_1 = 0.9 \quad z_2 = 0.1$

$z_1 = 0.5 \quad z_2 = 0.5$

$z_1 \cdot z_2 = 0.09$

$\frac{1}{2^2} = \frac{1}{4} = 0.25$

$z_1 \cdot z_2 = 0.25$

$z_1 \frac{1}{k} \quad z_2 = \frac{1}{k} \quad z_3 = \frac{1}{k} \quad z_4 \ldots$

Example

Preparation for Max. Spacing Estimator. Given set sample $(x_1, \dots, x_n)$

Consider order statistic:

$$x_{(1)} \leq \dots \leq x_{(n)}$$

$x_{(0)} := -\infty$, $x_{(n+1)} = +\infty$

$\rightsquigarrow \theta$ is characterised by the so called spacings

(θ: unknown)

$$D_i(\theta) = F_\theta(x_{(i)}) - F_\theta(x_{(i-1)}) \quad i = 1, \dots, n+1$$

Note: $D_i(\theta)$ is a realisation of the spacings of n RVs that are distributed on (0,1)

$F_\theta(X_i)$

of the same size the more uniform $F_\theta(x_{(i)})$ $\forall i$ are spacings should be roughly

Maximum-spacing method

Lemma: The maximum-spacing method is defined via

$$\widehat{\theta}_{MS} = \arg\max_{\theta\in\Theta} \prod_{i=1}^{n+1} \left(F_\theta(x_{(i)}) - F_\theta(x_{(i-1)})\right) \tag{4}$$

$\mathbb{P}_\theta[\tilde{F}_{\tilde{\theta}}(X_i) \le x] = x$

$F_\theta(F_{\tilde{\theta}}^{\wedge})$

$p = 0.2 \rightsquigarrow x_1, \dots, x_n$

$\widehat{\theta} = \tilde{p} = 0.8$

Example

Example

Unbiased estimators

Estimator

Def:

- An estimator is an arbitrary (Borel-measurable) function

$$\widehat{\theta} : \mathcal{X} \to \Theta, \quad x \mapsto \widehat{\theta}(x) \tag{5}$$

- An estimator $\widehat{\theta}$ is called unbiased, if

$$\mathbb{E}_\theta[\widehat{\theta}(X)] = \theta \tag{6}$$

  for all $\theta \in \Theta$.

- The bias of an estimator $\widehat{\theta}$ is

$$\text{Bias}_\theta(\widehat{\theta}) = \mathbb{E}_\theta[\widehat{\theta}(X)] - \theta \tag{7}$$

Note: $\text{Bias}_\theta(\widehat{\theta})$ is a function in $\widehat{\theta}$

Example

Example

Example

Mean square error

Def: Let $\Theta = (a, b) \subset \mathbb{R}$ be an interval. The mean square error (MSE) of an estimator $\widehat{\theta} : \mathcal{X} \to \Theta$

$$\mathrm{MSE}_\theta(\widehat{\theta}) = \mathbb{E}_\theta[(\widehat{\theta}(X) - \theta)^2] \tag{8}$$

Mean square error

Lemma: The relationship between the mean square error (MSE) of an estimator $\widehat{\theta} : \mathcal{X} \to \Theta$ and the BIAS is given by

$$\mathrm{MSE}_\theta(\widehat{\theta}) = \mathrm{Var}_\theta \widehat{\theta} + (\mathrm{Bias}_\theta(\widehat{\theta}))^2 \tag{9}$$

$\mathrm{MSE}_\theta(\widehat{\theta}) = \mathrm{Var}_\theta \widehat{\theta}$ 0

Proof

Proof

Consistently better

Let $\widehat{\Theta}_1$ and $\widehat{\Theta}_2$ be two estimators. The estimator θ_1 is consistently better than θ_2 if,

$$MSE_\theta(\widehat{\theta}_1) \leq MSE_\theta(\widehat{\theta}_2) \quad \forall \theta \in \Theta \tag{10}$$

Minimum-variance unbiased estimator

Def: An unbiased estimator $\widehat{\theta}$ is called minimum-variance unbiased estimator if all unbiased estimators $\tilde{\theta}$ the following inequality holds

$$\mathrm{Var}_\theta \widehat{\theta} \leq \mathrm{Var}_\theta \tilde{\theta} \tag{11}$$

for all $\theta \in \Theta$.

Minimum-variance unbiased estimator

Lemma: Let $\widehat{\theta}_1, \widehat{\theta}_2 : \mathcal{X} \to \Theta$ are two minimum-variance unbiased estimator the

$$\widehat{\theta}_1 = \widehat{\theta}_2 \quad \text{almost surely under } \mathbb{P} \text{ for all } \theta \in \Theta \tag{12}$$

for all $\theta \in \Theta$.

Proof

Proof

Proof

Sufficient statistic

Def: A function $T : \mathcal{X} \to \mathbb{R}^r$ is called a sufficient statistic if the function

$$\theta \mapsto \mathbb{P}_\theta[X = x | T(X) = t] \tag{13}$$

is constant for all $x \in \mathcal{X}$ and for all $t \in \mathbb{R}^r$,i.e.,

$$\mathbb{P}_{\theta_1}[X = x | T(X) = t]\mathbb{P}_{\theta_2}[X = x | T(X) = t] \tag{14}$$

for all $t \in \mathbb{R}^r$ and all $\theta_1, \theta_2 \in \Theta$ with $\mathbb{P}_{\theta_1}[T(X) = t] \neq 0$ and $\mathrm{P}_{\theta_2}[T(X) = t] \neq 0$