Assimilating data with outer probability measures

Jérémie Houssineau

Department of Statistics
The University of Warwick

April 9, 2019

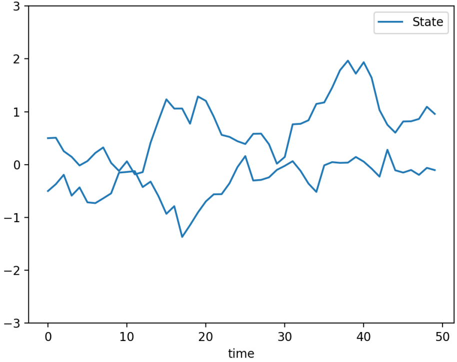
Joint work with: Emmanuel Delande (Texas Uni), Branko Ristić (RMIT), Neil Chada (NUS) and Adrian Bishop (UTS/CSIRO).

Outline

Outline

Multi-target tracking

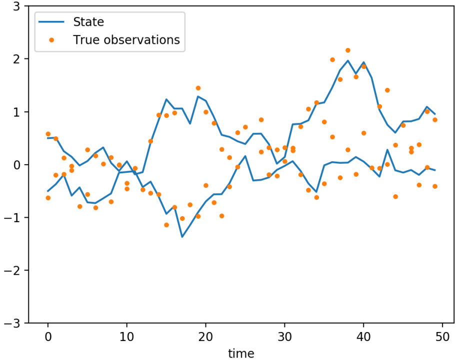
Consider two target trajectories.



Simple scenario

Multi-target tracking

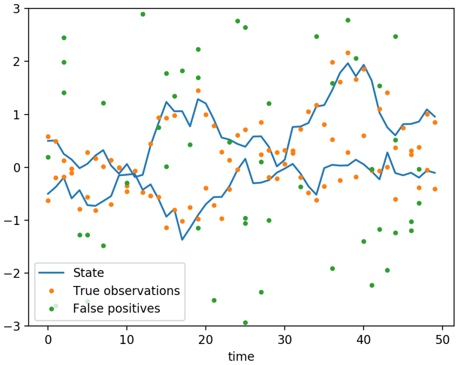
Consider two target trajectories.



Simple scenario

Multi-target tracking

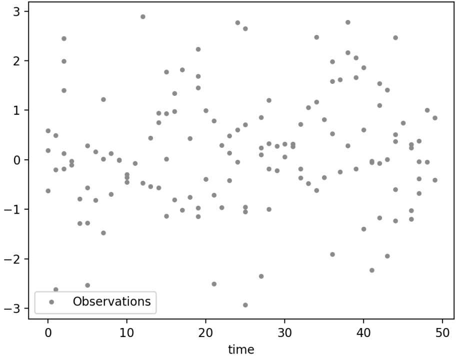
Consider two target trajectories.



Simple scenario

Multi-target tracking

Consider two target trajectories.



Simple scenario

Modelling

- Number of objects changes in time (birth/death process)
- Observation process:
  - FN Observation of a given object might fail
  - TP When successful, it is prone to errors $\rightarrow \ell(y \mid x)$
  - FP Some observations originate from background noise
- Data association is unknown a priori
- States and observations are represented by a point process (random counting measure)

$$\mathcal{X}_n = \sum_{i=1}^{N} \delta_{X_n^i} \qquad \text{and} \qquad \mathcal{Y}_n = \sum_{i=1}^{M} \delta_{Y_n^i}$$

- Interest: first-moment of the state point process (intensity measure)

$$F_n(B) = \mathbb{E}(\mathcal{X}_n(B)) = \mathbb{E}\left(\sum_{i=1}^{N} \mathbf{1}_B(X_n^i)\right)$$

Modelling

- Number of objects changes in time (birth/death process)
- Observation process:
  - FN Observation of a given object might fail
  - TP When successful, it is prone to errors $\rightarrow \ell(y \,|\, x)$
  - FP Some observations originate from background noise
- Data association is unknown a priori
- States and observations are represented by a point process (random counting measure)

$$\mathcal{X}_n = \sum_{i=1}^{N} \delta_{X_n^i} \qquad \text{and} \qquad \mathcal{Y}_n = \sum_{i=1}^{M} \delta_{Y_n^i}$$

- Interest: first-moment of the state point process (intensity measure)

$$F_n(B) = \mathbb{E}(\mathcal{X}_n(B)) = \mathbb{E}\left(\sum_{i=1}^{N} \mathbf{1}_B(X_n^i)\right)$$

Modelling

- Number of objects changes in time (birth/death process)
- Observation process:
  - FN Observation of a given object might fail
  - TP When successful, it is prone to errors $\rightarrow \ell(y \,|\, x)$
  - FP Some observations originate from background noise
- Data association is unknown a priori
- States and observations are represented by a point process (random counting measure)

$$\mathcal{X}_n = \sum_{i=1}^{N} \delta_{X_n^i} \qquad \text{and} \qquad \mathcal{Y}_n = \sum_{i=1}^{M} \delta_{Y_n^i}$$

- Interest: first-moment of the state point process (intensity measure)

$$F_n(B) = \mathbb{E}(\mathcal{X}_n(B)) = \mathbb{E}\left(\sum_{i=1}^{N} \mathbf{1}_B(X_n^i)\right)$$

Modelling

- Number of objects changes in time (birth/death process)
- Observation process:
  - FN Observation of a given object might fail
  - TP When successful, it is prone to errors $\rightarrow \ell(y \,|\, x)$
  - FP Some observations originate from background noise
- Data association is unknown a priori
- States and observations are represented by a point process (random counting measure)

$$\mathcal{X}_n = \sum_{i=1}^{N} \delta_{X_n^i} \qquad \text{and} \qquad \mathcal{Y}_n = \sum_{i=1}^{M} \delta_{Y_n^i}$$

- Interest: first-moment of the state point process (intensity measure)

$$F_n(B) = \mathbb{E}(\mathcal{X}_n(B)) = \mathbb{E}\Big(\sum_{i=1}^{N} \mathbf{1}_B(X_n^i)\Big)$$

Modelling

- Number of objects changes in time (birth/death process)
- Observation process:
  - FN Observation of a given object might fail
  - TP When successful, it is prone to errors $\to \ell(y \,|\, x)$
  - FP Some observations originate from background noise
- Data association is unknown a priori
- States and observations are represented by a point process (random counting measure)

$$\mathcal{X}_n = \sum_{i=1}^{N} \delta_{X_n^i} \qquad \text{and} \qquad \mathcal{Y}_n = \sum_{i=1}^{M} \delta_{Y_n^i}$$

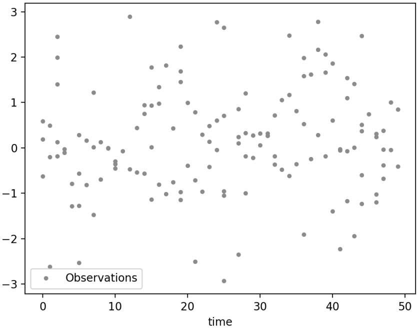
- Interest: first-moment of the state point process (intensity measure)

$$F_n(B) = \mathbb{E}(\mathcal{X}_n(B)) = \mathbb{E}\left(\sum_{i=1}^{N} \mathbf{1}_B(X_n^i) \right)$$

PHD filter

Update equation at time step n

$$F_n(x \,|\, \mathcal{Y}_{1:n}) = p_{\mathrm{FN}}\, F_n(x \,|\, \mathcal{Y}_{1:n-1}) + \sum_{y \in \mathcal{Y}_n} \frac{p_{\mathrm{TP}} \ell(y \,|\, x) F_n(x \,|\, \mathcal{Y}_{1:n-1})}{F_{\mathrm{FP}}(y) + p_{\mathrm{TP}} \int \ell(y \,|\, x') F_n(x' \,|\, \mathcal{Y}_{1:n-1}) \mathrm{d}x'}$$

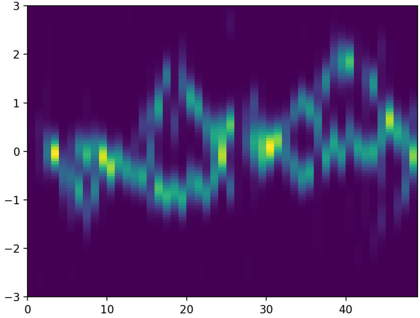


R.P.S. Mahler. "Multitarget Bayes filtering via first-order multitarget moments". In: *IEEE Transactions on Aerospace and Electronic systems* 39.4 (2003)

PHD filter

Update equation at time step n

$$F_n(x \mid \mathcal{Y}_{1:n}) = p_{\mathrm{FN}}\, F_n(x \mid \mathcal{Y}_{1:n-1}) + \sum_{y \in \mathcal{Y}_n} \frac{p_{\mathrm{TP}} \ell(y \mid x) F_n(x \mid \mathcal{Y}_{1:n-1})}{F_{\mathrm{FP}}(y) + p_{\mathrm{TP}} \int \ell(y \mid x') F_n(x' \mid \mathcal{Y}_{1:n-1}) \mathrm{d}x'}$$

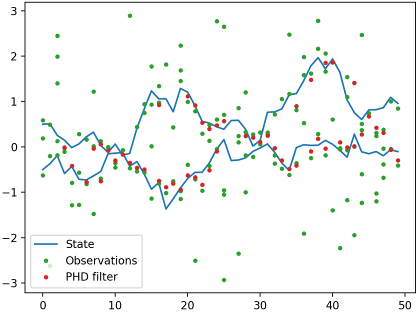


R.P.S. Mahler. "Multitarget Bayes filtering via first-order multitarget moments". In: *IEEE Transactions on Aerospace and Electronic systems* 39.4 (2003)

PHD filter

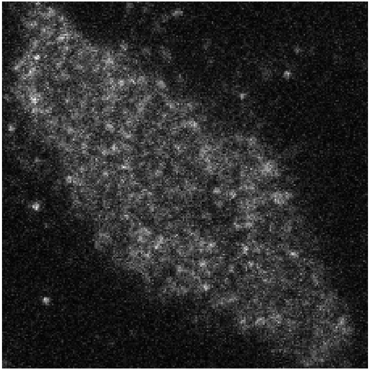
Update equation at time step n

$$F_n(x \,|\, \mathcal{Y}_{1:n}) = p_{\mathrm{FN}}\, F_n(x \,|\, \mathcal{Y}_{1:n-1}) + \sum_{y \in \mathcal{Y}_n} \frac{p_{\mathrm{TP}} \ell(y \,|\, x) F_n(x \,|\, \mathcal{Y}_{1:n-1})}{F_{\mathrm{FP}}(y) + p_{\mathrm{TP}} \int \ell(y \,|\, x') F_n(x' \,|\, \mathcal{Y}_{1:n-1}) \mathrm{d}x'}$$

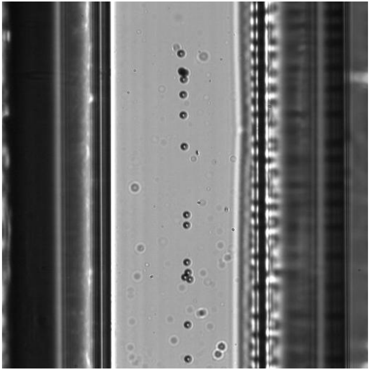


R.P.S. Mahler. "Multitarget Bayes filtering via first-order multitarget moments". In: *IEEE Transactions on Aerospace and Electronic systems* 39.4 (2003)

Application: Life sciences

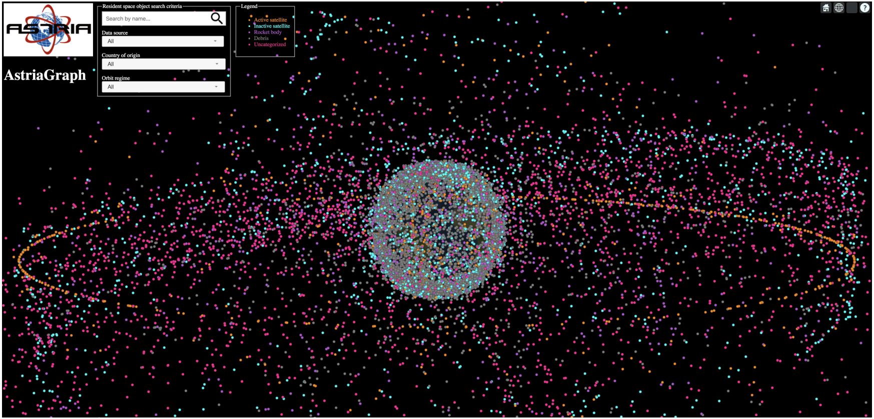


(a) Fluorescence microscopy



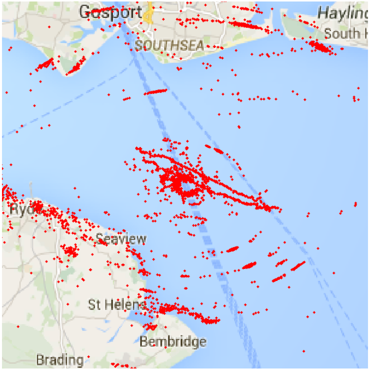
(b) Microfluidics for water safety

Application: Space Situational Awareness



http://astria.tacc.utexas.edu/AstriaGraph/

Application: Surveillance and defence



(a) Radar data from a ship

(b) Surveillance of a harbour area

Fundamental limitations

- The distribution of false positives can vary dramatically in time
- There is often no prior information on the location of objects
- The observation process is difficult to describe in a standard way

Fundamental limitations

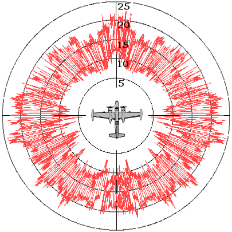
- The distribution of false positives can vary dramatically in time
- There is often no prior information on the location of objects
- The observation process is difficult to describe in a standard way

Fundamental limitations

- The distribution of false positives can vary dramatically in time
- There is often no prior information on the location of objects
- The observation process is difficult to describe in a standard way

Fundamental limitations

- The distribution of false positives can vary dramatically in time
- There is often no prior information on the location of objects
- The observation process is difficult to describe in a standard way



Radar cross section of an A-26 Invader (Wikipedia)

Outline

Outer probability measure

Assuming:

- A r.v. X on $\mathbf{X}$ with conditional probability distribution $p(\cdot \,|\, \theta)$
- No knowledge about $\theta \in \Theta$

Then

$$\mathbb{P}(X \in B) \leq \sup_{\theta \in \Theta} p(B \,|\, \theta)$$

With $f : \Theta \to [0,1]$ such that $\sup_\theta f(\theta) = 1$

Definition

The set function $\bar{P}$ in $[0,1]$ is an outer probability measure since

1. $\bar{P}(\emptyset) = 0$ and $\bar{P}(\mathbf{X}) = 1$
2. monotone: $A \subseteq B$ implies $\bar{P}(A) \leq \bar{P}(B)$
3. sub-additive: $\bar{P}(A \cup B) \leq \bar{P}(A) + \bar{P}(B)$ for any $A, B \subseteq \mathbf{X}$

J. H. "Parameter estimation with a class of outer probability measures". In: *arXiv:1801.00569* (2018)

Outer probability measure

Assuming:

- A r.v. X on $\mathbf{X}$ with conditional probability distribution $p(\cdot \,|\, \theta)$
- No knowledge about $\theta \in \Theta$

Then

$$\mathbb{P}(X \in B) \leq \sup_{\theta \in \Theta} p(B \,|\, \theta) \mathbf{1}_A(\theta)$$

With $f : \Theta \to [0,1]$ such that $\sup_\theta f(\theta) = 1$

Definition

The set function $\bar{P}$ in $[0,1]$ is an outer probability measure since

1. $\bar{P}(\emptyset) = 0$ and $\bar{P}(\mathbf{X}) = 1$
2. monotone: $A \subseteq B$ implies $\bar{P}(A) \leq \bar{P}(B)$
3. sub-additive: $\bar{P}(A \cup B) \leq \bar{P}(A) + \bar{P}(B)$ for any $A, B \subseteq \mathbf{X}$

J. H. "Parameter estimation with a class of outer probability measures". In: *arXiv:1801.00569* (2018)

Outer probability measure

Assuming:

- A r.v. X on $\mathbf{X}$ with conditional probability distribution $p(\cdot \,|\, \theta)$
- No knowledge about $\theta \in \Theta$

Then

$$\mathbb{P}(X \in B) \leq \sup_{\theta \in \Theta} p(B \,|\, \theta) f(\theta)$$

With $f : \Theta \to [0, 1]$ such that $\sup_\theta f(\theta) = 1$

Definition

The set function $\bar{P}$ in $[0, 1]$ is an outer probability measure since

1. $\bar{P}(\emptyset) = 0$ and $\bar{P}(\mathbf{X}) = 1$
2. monotone: $A \subseteq B$ implies $\bar{P}(A) \leq \bar{P}(B)$
3. sub-additive: $\bar{P}(A \cup B) \leq \bar{P}(A) + \bar{P}(B)$ for any $A, B \subseteq \mathbf{X}$

J. H. "Parameter estimation with a class of outer probability measures". In: *arXiv:1801.00569* (2018)

Outer probability measure

Assuming:

- A r.v. X on $\mathbf{X}$ with conditional probability distribution $p(\cdot \,|\, \theta)$
- No knowledge about $\theta \in \Theta$

Then

$$\mathbb{P}(X \in B) \leq \sup_{\theta \in \Theta} p(B \,|\, \theta) f(\theta)$$

With $f : \Theta \to [0, 1]$ such that $\sup_\theta f(\theta) = 1$

Definition

The set function $\bar{P}$ in $[0, 1]$ is an outer probability measure since

1. $\bar{P}(\emptyset) = 0$ and $\bar{P}(\mathbf{X}) = 1$
2. monotone: $A \subseteq B$ implies $\bar{P}(A) \leq \bar{P}(B)$
3. sub-additive: $\bar{P}(A \cup B) \leq \bar{P}(A) + \bar{P}(B)$ for any $A, B \subseteq \mathbf{X}$

J. H. "Parameter estimation with a class of outer probability measures". In: *arXiv:1801.00569* (2018)

Outer probability measure

Assuming:

- A r.v. X on $\mathbf{X}$ with conditional probability distribution $p(\cdot \,|\, \theta)$
- No knowledge about $\theta \in \Theta$

Then

$$\mathbb{P}(X \in B) \leq \sup_{\theta \in \Theta} p(B \,|\, \theta) f(\theta) = \bar{P}(B)$$

With $f : \Theta \to [0,1]$ such that $\sup_\theta f(\theta) = 1$

Definition

The set function $\bar{P}$ in $[0,1]$ is an outer probability measure since

1. $\bar{P}(\emptyset) = 0$ and $\bar{P}(\mathbf{X}) = 1$
2. monotone: $A \subseteq B$ implies $\bar{P}(A) \leq \bar{P}(B)$
3. sub-additive: $\bar{P}(A \cup B) \leq \bar{P}(A) + \bar{P}(B)$ for any $A, B \subseteq \mathbf{X}$

J. H. “Parameter estimation with a class of outer probability measures”. In: *arXiv:1801.00569* (2018)

Outer probability measure

Assuming:

- A r.v. X on $\mathbf{X}$ with conditional probability distribution $p(\cdot \,|\, \theta)$
- No knowledge about $\theta \in \Theta$

Then

$$\mathbb{P}(X \in B) \leq \sup_{\theta \in \Theta} p(B \,|\, \theta) f(\theta) = \bar{P}(B)$$

With $f : \Theta \to [0, 1]$ such that $\sup_\theta f(\theta) = 1$

Definition

The set function $\bar{P}$ in $[0, 1]$ is an outer probability measure since

1. $\bar{P}(\emptyset) = 0$ and $\bar{P}(\mathbf{X}) = 1$
2. monotone: $A \subseteq B$ implies $\bar{P}(A) \leq \bar{P}(B)$
3. sub-additive: $\bar{P}(A \cup B) \leq \bar{P}(A) + \bar{P}(B)$ for any $A, B \subseteq \mathbf{X}$

J. H. "Parameter estimation with a class of outer probability measures". In: *arXiv:1801.00569* (2018)

Possibility function

	Parameter(s)	Function of $x \in \mathbb{R}$
Uniform $\bar{\mathcal{U}}([a,b])$	$a, b \in \mathbb{R}$, $a < b$	$\mathbf{1}_{[a,b]}(x)$
Gaussian $\bar{\mathcal{N}}(\mu, \sigma^2)$	$\mu \in \mathbb{R}$, $\sigma^2 > 0$	$\exp\left(-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right)$
Student's t	$\nu > 0$	$\left(1 + \frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}}$
Cauchy	$x_0 \in \mathbb{R}$, $\gamma > 0$	$\frac{\gamma^2}{(x-x_0)^2 + \gamma^2}$
Gamma	$k \geq 0$, $\theta > 0$	$\left(\frac{x}{k\theta}\right)^k \exp\left(-\frac{x}{\theta} + k\right)$

Possibility function

	Parameter(s)	Function of $x \in \mathbb{R}$
Uniform $\bar{\mathcal{U}}([a,b])$	$a, b \in \mathbb{R},\ a < b$	$\mathbf{1}_{[a,b]}(x)$
Gaussian $\bar{\mathcal{N}}(\mu, \sigma^2)$	$\mu \in \mathbb{R},\ \sigma^2 > 0$	$\exp\Big(-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\Big)$
Student's t	$\nu > 0$	$\frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\pi}\Gamma(\frac{\nu}{2})}\Big(1+\frac{x^2}{\nu}\Big)^{-\frac{\nu+1}{2}}$
Cauchy	$x_0 \in \mathbb{R},\ \gamma > 0$	$\frac{\gamma^2}{(x-x_0)^2+\gamma^2}$
Gamma	$k \geq 0,\ \theta > 0$	$\Big(\frac{x}{k\theta}\Big)^k \exp\Big(-\frac{x}{\theta}+k\Big)$

Uncertain variable

Ingredients:

- A sample space Ω_{u} for deterministic but uncertainty phenomena
- A probability space $(\Omega_{\mathrm{r}}, \mathcal{F}, \mathbb{P}(\cdot \,|\, \omega_{\mathrm{u}}))$ for any $\omega_{\mathrm{u}} \in \Omega_{\mathrm{u}}$
- A state space $\mathbf{X}$ and a parameter space Θ

Definition

An uncertain variable (u.v.) is a mapping

$$\begin{aligned} X : \Omega_{\mathrm{u}} \times \Omega_{\mathrm{r}} &\to \Theta \times \mathbf{X} \\ (\omega_{\mathrm{u}}, \omega_{\mathrm{r}}) &\mapsto (X_{\mathrm{u}}(\omega_{\mathrm{u}}), X_{\mathrm{r}}(\omega_r)) \end{aligned}$$

such that

- $X_{\mathrm{r}} : \Omega_{\mathrm{r}} \to \mathbf{X}$ is a random variable
- $\mathbb{P}(X_{\mathrm{r}}^{-1}(B) \,|\, \cdot)$ is constant over $X_{\mathrm{u}}^{-1}[\theta]$ for any $B \subseteq \mathbf{X}$ and $\theta \in \Theta$
  1. implies that θ is sufficiently informative about X_r
  2. can deduce the conditional distribution $p(\cdot \,|\, \theta)$

Uncertain variable

Ingredients:

- A sample space Ω_u for deterministic but uncertainty phenomena
- A probability space $(\Omega_\mathrm{r}, \mathcal{F}, \mathbb{P}(\cdot \,|\, \omega_\mathrm{u}))$ for any $\omega_\mathrm{u} \in \Omega_\mathrm{u}$
- A state space $\mathbf{X}$ and a parameter space Θ

Definition

An uncertain variable (u.v.) is a mapping

$$\begin{aligned} X : \Omega_\mathrm{u} \times \Omega_\mathrm{r} &\to \Theta \times \mathbf{X} \\ (\omega_\mathrm{u}, \omega_\mathrm{r}) &\mapsto (X_\mathrm{u}(\omega_\mathrm{u}), X_\mathrm{r}(\omega_r)) \end{aligned}$$

such that

- $X_\mathrm{r} : \Omega_\mathrm{r} \to \mathbf{X}$ is a random variable
- $\mathbb{P}(X_\mathrm{r}^{-1}(B) \,|\, \cdot)$ is constant over $X_\mathrm{u}^{-1}[\theta]$ for any $B \subseteq \mathbf{X}$ and $\theta \in \Theta$
  1. implies that θ is sufficiently informative about X_r
  2. can deduce the conditional distribution $p(\cdot \,|\, \theta)$

Uncertain variable

Ingredients:

- A sample space Ω_{u} for deterministic but uncertainty phenomena
- A probability space $(\Omega_{\mathrm{r}}, \mathcal{F}, \mathbb{P}(\cdot \,|\, \omega_{\mathrm{u}}))$ for any $\omega_{\mathrm{u}} \in \Omega_{\mathrm{u}}$
- A state space $\mathbf{X}$ and a parameter space Θ

Definition

An uncertain variable (u.v.) is a mapping

$$\begin{aligned} X : \Omega_{\mathrm{u}} \times \Omega_{\mathrm{r}} &\to \Theta \times \mathbf{X} \\ (\omega_{\mathrm{u}}, \omega_{\mathrm{r}}) &\mapsto (X_{\mathrm{u}}(\omega_{\mathrm{u}}), X_{\mathrm{r}}(\omega_r)) \end{aligned}$$

such that

- $X_{\mathrm{r}} : \Omega_{\mathrm{r}} \to \mathbf{X}$ is a random variable
- $\mathbb{P}(X_{\mathrm{r}}^{-1}(B) \,|\, \cdot)$ is constant over $X_{\mathrm{u}}^{-1}[\theta]$ for any $B \subseteq \mathbf{X}$ and $\theta \in \Theta$
    1. implies that θ is sufficiently informative about X_r
    2. can deduce the conditional distribution $p(\cdot \,|\, \theta)$

Assumption & basic concepts

Assumption

Henceforth: $p(\cdot \mid \theta) = \delta_\theta$ and $\Theta = \mathbf{X}$

Concept

The (deterministic) u.v.s X and Y are (weakly) independent if

$$f_{X,Y}(x, y) = f_X(x) f_Y(y)$$

Even if X and Y are not independent $\sqrt{f_X \times f_Y}$ is a valid description of (X, Y) with

$$f_X(x) = \sup_y f_{X,Y}(x, y) \qquad \text{and} \qquad f_Y(y) = \sup_x f_{X,Y}(x, y)$$

$\rightarrow$ information loss

Assumption & basic concepts

Assumption

Henceforth: $p(\cdot \mid \theta) = \delta_\theta$ and $\Theta = \mathbf{X}$

Concept

The (deterministic) u.v.s X and Y are (weakly) independent if

$$f_{X,Y}(x, y) = f_X(x) f_Y(y)$$

Even if X and Y are not independent $\sqrt{f_X \times f_Y}$ is a valid description of (X, Y) with

$$f_X(x) = \sup_y f_{X,Y}(x, y) \qquad \text{and} \qquad f_Y(y) = \sup_x f_{X,Y}(x, y)$$

$\rightarrow$ information loss

Assumption & basic concepts

Assumption

Henceforth: $p(\cdot \,|\, \theta) = \delta_\theta$ and $\Theta = \mathbf{X}$

Concept

The (deterministic) u.v.s X and Y are (weakly) independent if

$$f_{X,Y}(x, y) = f_X(x) f_Y(y)$$

Even if X and Y are not independent $\sqrt{f_X \times f_Y}$ is a valid description of (X, Y) with

$$f_X(x) = \sup_y f_{X,Y}(x, y) \qquad \text{and} \qquad f_Y(y) = \sup_x f_{X,Y}(x, y)$$

$\rightarrow$ information loss

Assumption & basic concepts

Assumption

Henceforth: $p(\cdot \,|\, \theta) = \delta_\theta$ and $\Theta = \mathbf{X}$

Concept

The (deterministic) u.v.s X and Y are (weakly) independent if

$$f_{X,Y}(x, y) = f_X(x) f_Y(y)$$

Even if X and Y are not independent $\sqrt{f_X \times f_Y}$ is a valid description of (X, Y) with

$$f_X(x) = \sup_y f_{X,Y}(x, y) \qquad \text{and} \qquad f_Y(y) = \sup_x f_{X,Y}(x, y)$$

$\rightarrow$ information loss

Possibility function

Pros & cons

- Does not require a reference measure
- Standard operations apply directly: if (X, Y) u.v. described by $f_{X,Y}$

$$f_Y(y) = \sup_x f_{X,Y}(x, y) \qquad \text{and} \qquad f_{X|Y}(x \,|\, y) = \frac{f_{X,Y}(x, y)}{f_Y(y)}$$

- Can be truncated, discretized
- Straightforward to introduce new possibility functions
- Easy to extent existing concepts, e.g. a collection of u.v.s $\{X_n\}_n$ is a (weak) Markov chain if

$$f_{X_n}(\cdot \,|\, X_{1:n-1}) = f_{X_n}(\cdot \,|\, X_{n-1})$$

- Less obvious for distribution on $\mathbb{N}$
- Cannot sample

Possibility function

Pros & cons

- Does not require a reference measure
- Standard operations apply directly: if (X, Y) u.v. described by $f_{X,Y}$

$$f_Y(y) = \sup_x f_{X,Y}(x, y) \qquad \text{and} \qquad f_{X|Y}(x \,|\, y) = \frac{f_{X,Y}(x, y)}{f_Y(y)}$$

- Can be truncated, discretized
- Straightforward to introduce new possibility functions
- Easy to extent existing concepts, e.g. a collection of u.v.s $\{X_n\}_n$ is a (weak) Markov chain if

$$f_{X_n}(\cdot \,|\, X_{1:n-1}) = f_{X_n}(\cdot \,|\, X_{n-1})$$

- Less obvious for distribution on $\mathbb{N}$
- Cannot sample

Possibility function

Pros & cons

- Does not require a reference measure
- Standard operations apply directly: if (X, Y) u.v. described by $f_{X,Y}$

$$f_Y(y) = \sup_x f_{X,Y}(x, y) \qquad \text{and} \qquad f_{X|Y}(x \mid y) = \frac{f_{X,Y}(x, y)}{f_Y(y)}$$

- Can be truncated, discretized
- Straightforward to introduce new possibility functions
- Easy to extent existing concepts, e.g. a collection of u.v.s $\{X_n\}_n$ is a (weak) Markov chain if

$$f_{X_n}(\cdot \mid X_{1:n-1}) = f_{X_n}(\cdot \mid X_{n-1})$$

- Less obvious for distribution on $\mathbb{N}$
- Cannot sample

Possibility function

Pros & cons

- Does not require a reference measure
- Standard operations apply directly: if (X, Y) u.v. described by $f_{X,Y}$

$$f_Y(y) = \sup_x f_{X,Y}(x, y) \qquad \text{and} \qquad f_{X|Y}(x \,|\, y) = \frac{f_{X,Y}(x, y)}{f_Y(y)}$$

- Can be truncated, discretized
- Straightforward to introduce new possibility functions
- Easy to extent existing concepts, e.g. a collection of u.v.s $\{X_n\}_n$ is a (weak) Markov chain if

$$f_{X_n}(\cdot \,|\, X_{1:n-1}) = f_{X_n}(\cdot \,|\, X_{n-1})$$

- Less obvious for distribution on $\mathbb{N}$
- Cannot sample

Possibility function

Pros & cons

- Does not require a reference measure
- Standard operations apply directly: if (X, Y) u.v. described by $f_{X,Y}$

$$f_Y(y) = \sup_x f_{X,Y}(x, y) \qquad \text{and} \qquad f_{X|Y}(x \,|\, y) = \frac{f_{X,Y}(x, y)}{f_Y(y)}$$

- Can be truncated, discretized
- Straightforward to introduce new possibility functions
- Easy to extent existing concepts, e.g. a collection of u.v.s $\{X_n\}_n$ is a (weak) Markov chain if

$$f_{X_n}(\cdot \,|\, X_{1:n-1}) = f_{X_n}(\cdot \,|\, X_{n-1})$$

- Less obvious for distribution on $\mathbb{N}$
- Cannot sample

Possibility function

Pros & cons

- Does not require a reference measure
- Standard operations apply directly: if (X, Y) u.v. described by $f_{X,Y}$

$$f_Y(y) = \sup_x f_{X,Y}(x, y) \qquad \text{and} \qquad f_{X|Y}(x \,|\, y) = \frac{f_{X,Y}(x, y)}{f_Y(y)}$$

- Can be truncated, discretized
- Straightforward to introduce new possibility functions
- Easy to extent existing concepts, e.g. a collection of u.v.s $\{X_n\}_n$ is a (weak) Markov chain if

$$f_{X_n}(\cdot \,|\, X_{1:n-1}) = f_{X_n}(\cdot \,|\, X_{n-1})$$

- Less obvious for distribution on $\mathbb{N}$
- Cannot sample

Possibility function

Pros & cons

- Does not require a reference measure
- Standard operations apply directly: if (X, Y) u.v. described by $f_{X,Y}$

$$f_Y(y) = \sup_x f_{X,Y}(x, y) \qquad \text{and} \qquad f_{X|Y}(x \,|\, y) = \frac{f_{X,Y}(x, y)}{f_Y(y)}$$

- Can be truncated, discretized
- Straightforward to introduce new possibility functions
- Easy to extent existing concepts, e.g. a collection of u.v.s $\{X_n\}_n$ is a (weak) Markov chain if

$$f_{X_n}(\cdot \,|\, X_{1:n-1}) = f_{X_n}(\cdot \,|\, X_{n-1})$$

- Less obvious for distribution on $\mathbb{N}$
- Cannot sample

Expectation

By identification

For a given function φ

$$\bar{\mathbb{E}}_f(\varphi(X)) = \sup_{x \in \mathbf{X}} \varphi(x) f(x)$$

Intuitively

$$\mathbb{E}^*_f(X) = \operatorname*{argmax}_x f(x)$$

Expectation

By identification

For a given function φ

$$\bar{\mathbb{E}}_f(\varphi(X)) = \sup_{x \in \mathbf{X}} \varphi(x) f(x)$$

Example: Define the self-information as $\bar{\mathrm{I}}(x) = -\log f(x)$ then

$$\bar{\mathrm{H}}(X) = \bar{\mathbb{E}}_f(\bar{\mathrm{I}}(X)) = \sup_{x \in \mathbf{X}} \big(- f(x) \log f(x) \big)$$

$\rightarrow$ meaningful on uncountable spaces

Intuitively

$$\mathbb{E}^*_f(X) = \operatorname*{argmax}_x f(x)$$

Expectation

By identification

For a given function φ

$$\bar{\mathbb{E}}_f(\varphi(X)) = \sup_{x \in \mathbf{X}} \varphi(x) f(x)$$

Intuitively

$$\mathbb{E}^*_f(X) = \operatorname*{argmax}_x f(x)$$

Expectation

By identification

For a given function φ

$$\bar{\mathbb{E}}_f(\varphi(X)) = \sup_{x \in \mathbf{X}} \varphi(x) f(x)$$

Intuitively

$$\mathbb{E}^*_f(X) = \operatorname*{argmax}_x f(x)$$

Example: Maximum-likelihood estimate with i.i.d. samples $y_1, \ldots, y_n \sim p(\cdot \,|\, x)$

$$\mathbb{E}^*_f(X \,|\, y_{1:n}) = \operatorname*{argmax}_x f(x) \prod_{i=1}^{n} p(y_i \,|\, x)$$

$\rightarrow$ can justify profile likelihood

Law of large numbers

Statement

Consider:

- Some u.v.s $X_1, X_2, \dots$ on $\mathbb{R}^d$ i.i.d. by f
- The u.v. $S_n = n^{-1} \sum_{i=1}^{n} X_n$ described by

$$f_{S_n}(y) = \sup\Big\{ \prod_{i=1}^{n} f(x_i) : \frac{1}{n}(x_1 + \dots + x_n) = y \Big\}.$$

Proposition

Let f be a possibility function such that $f(x) \to 0$ when $\|x\| \to \infty$ and $\operatorname{argmax}_x f(x) = \mu$. If f verifies some additional regularity conditions then it holds that

$$\lim_{n\to\infty} f_{S_n} = \mathbf{1}_{\mu},$$

where the limit is considered point-wise.

$\rightarrow$ Confirms the intuitive definition of expectation

Law of large numbers

Statement

Consider:

- Some u.v.s $X_1, X_2, \dots$ on $\mathbb{R}^d$ i.i.d. by f
- The u.v. $S_n = n^{-1} \sum_{i=1}^{n} X_n$ described by

$$f_{S_n}(y) = \sup \Big\{ \prod_{i=1}^{n} f(x_i) : \frac{1}{n}(x_1 + \dots + x_n) = y \Big\}.$$

Proposition

Let f be a possibility function such that $f(x) \to 0$ when $\|x\| \to \infty$ and $\operatorname{argmax}_x f(x) = \mu$. If f verifies some additional regularity conditions then it holds that

$$\lim_{n \to \infty} f_{S_n} = \mathbf{1}_{\mu},$$

where the limit is considered point-wise.

$\rightarrow$ Confirms the intuitive definition of expectation

Law of large numbers

Statement

Consider:

- Some u.v.s $X_1, X_2, \ldots$ on $\mathbb{R}^d$ i.i.d. by f
- The u.v. $S_n = n^{-1} \sum_{i=1}^n X_n$ described by

$$f_{S_n}(y) = \sup \bigg\{ \prod_{i=1}^n f(x_i) : \frac{1}{n}(x_1 + \cdots + x_n) = y \bigg\}.$$

Proposition

Let f be a possibility function such that $f(x) \to 0$ when $\|x\| \to \infty$ and $\operatorname{argmax}_x f(x) = \mu$. If f verifies some additional regularity conditions then it holds that

$$\lim_{n\to\infty} f_{S_n} = \mathbf{1}_\mu,$$

where the limit is considered point-wise.

$\to$ Confirms the intuitive definition of expectation

Law of large numbers

Statement

Consider:

- Some u.v.s $X_1, X_2, \dots$ on $\mathbb{R}^d$ i.i.d. by f
- The u.v. $S_n = n^{-1} \sum_{i=1}^n X_n$ described by

$$f_{S_n}(y) = \sup \Big\{ \prod_{i=1}^n f(x_i) : \frac{1}{n}(x_1 + \cdots + x_n) = y \Big\}.$$

Proposition

Let f be a possibility function such that $f(x) \to 0$ when $\|x\| \to \infty$ and $\operatorname{argmax}_x f(x) = \mu$. *If f verifies some additional regularity conditions then it holds that*

$$\lim_{n\to\infty} f_{S_n} = \mathbf{1}_\mu,$$

where the limit is considered point-wise.

$\to$ Confirms the intuitive definition of expectation

Law of large numbers

Example

1. Consider observations $Y, Y_1, Y_2, \dots$ i.i.d. by $f_Y(\cdot|\theta) = \bar{\mathcal{N}}(\cdot\,; \theta, \sigma^2)$
2. The sample average $S_n = n^{-1}\sum_{i=1}^n Y_i$ is a sufficient statistics, i.e.

$$f_{Y_{1:n}}(y_1, \dots, y_n \,|\, \theta) = \prod_{i=1}^n f_Y(y_i \,|\, \theta)$$
$$= \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^n (y_i - s_n)^2\right)\exp\left(-\frac{n}{2\sigma^2}(\theta - s_n)^2\right)$$

3. Consider instead the likelihood

$$f_{S_n}(s \,|\, \theta) = \sup_{\substack{y_1,\dots,y_n \\ n^{-1}\sum_i y_i = s}} f_{Y_{1:n}}(y_1, \dots, y_n \,|\, \theta) = \exp\left(-\frac{n}{2\sigma^2}(\theta - s)^2\right)$$

$\to$ tends to $\mathbf{1}_\theta$ as n grows by the LLN

Law of large numbers

Example

1. Consider observations $Y, Y_1, Y_2, \dots$ i.i.d. by $f_Y(\cdot|\theta) = \bar{\mathcal{N}}(\cdot\,; \theta, \sigma^2)$
2. The sample average $S_n = n^{-1}\sum_{i=1}^n Y_i$ is a sufficient statistics, i.e.

$$
\begin{aligned}
f_{Y_{1:n}}(y_1, \dots, y_n \,|\, \theta) &= \prod_{i=1}^n f_Y(y_i \,|\, \theta) \\
&= \exp\Big(-\frac{1}{2\sigma^2}\sum_{i=1}^n (y_i - s_n)^2\Big) \exp\Big(-\frac{n}{2\sigma^2}(\theta - s_n)^2\Big)
\end{aligned}
$$

3. Consider instead the likelihood

$$
f_{S_n}(s \,|\, \theta) = \sup_{\substack{y_1,\dots,y_n \\ n^{-1}\sum_i y_i = s}} f_{Y_{1:n}}(y_1, \dots, y_n \,|\, \theta) = \exp\Big(-\frac{n}{2\sigma^2}(\theta - s)^2\Big)
$$

$\to$ tends to $\mathbf{1}_\theta$ as n grows by the LLN

Law of large numbers

Example

1. Consider observations $Y, Y_1, Y_2, \dots$ i.i.d. by $f_Y(\cdot|\theta) = \bar{\mathcal{N}}(\cdot\,; \theta, \sigma^2)$
2. The sample average $S_n = n^{-1}\sum_{i=1}^n Y_i$ is a sufficient statistics, i.e.

$$
\begin{aligned}
f_{Y_{1:n}}(y_1, \dots, y_n \,|\, \theta) &= \prod_{i=1}^n f_Y(y_i \,|\, \theta) \\
&= \exp\Big(-\frac{1}{2\sigma^2}\sum_{i=1}^n (y_i - s_n)^2 \Big) \exp\Big(-\frac{n}{2\sigma^2}(\theta - s_n)^2 \Big)
\end{aligned}
$$

3. Consider instead the likelihood

$$
f_{S_n}(s \,|\, \theta) = \sup_{\substack{y_1, \dots, y_n \\ n^{-1}\sum_i y_i = s}} f_{Y_{1:n}}(y_1, \dots, y_n \,|\, \theta) = \exp\Big(-\frac{n}{2\sigma^2}(\theta - s)^2 \Big)
$$

$\to$ tends to $\mathbf{1}_\theta$ as n grows by the LLN

Central limit theorem

Statement

Proposition

If $\mathbb{E}^*_f(X) = \{\mu\}$ *and if* f *is strictly log-concave and twice differentiable at* μ*, then the possibility function* f_n *describing the u.v.* $\sqrt{n}(S_n - \mu)$ *verifies*

$$\lim_{n\to\infty} f_n(x) = \exp\Big(-\frac{1}{2}|f''(\mu)|(x-\mu)^2\Big)$$

for any x *in the domain of* f*.*

Consequences:

- Recover exactly the Laplace approximation
- Can study asymptotic properties of estimators like MLE

Central limit theorem

Statement

Proposition

If $\mathbb{E}^*_f(X) = \{\mu\}$ and if f is strictly log-concave and twice differentiable at μ, then the possibility function f_n describing the u.v. $\sqrt{n}(S_n - \mu)$ verifies

$$\lim_{n\to\infty} f_n(x) = \exp\Big(- \frac{1}{2}|f''(\mu)|(x-\mu)^2\Big)$$

for any x in the domain of f.

Consequences:

- Recover exactly the Laplace approximation
- Can study asymptotic properties of estimators like MLE

Central limit theorem

Variance

Variance of an u.v.

CLT suggests a definition of the variance as

$$\mathbb{V}^*_f(X) = -\frac{1}{f''(\mathbb{E}^*_f(X))}$$

Example (Gamma possibility function on $(0,\infty)$)

$$\mathbb{E}^*_f(X) = k\theta \qquad \text{and} \qquad \mathbb{V}^*_f(X) = k\theta^2,$$

with shape parameter $k \geq 0$ and scale parameter $\theta > 0$

Example (Cauchy possibility function on $\mathbb{R}$)

$$\mathbb{E}^*_f(X) = x_0 \qquad \text{and} \qquad \mathbb{V}^*_f(X) = \frac{\gamma^2}{2},$$

with location parameter $x_0 \in \mathbb{R}$ and scale parameter $\gamma > 0$

Central limit theorem

Variance

Variance of an u.v.

CLT suggests a definition of the variance as

$$\mathbb{V}^*_f(X) = -\frac{1}{f''(\mathbb{E}^*_f(X))}$$

Example (Gamma possibility function on $(0, \infty)$)

$$\mathbb{E}^*_f(X) = k\theta \qquad \text{and} \qquad \mathbb{V}^*_f(X) = k\theta^2,$$

with shape parameter $k \geq 0$ and scale parameter $\theta > 0$

Example (Cauchy possibility function on $\mathbb{R}$)

$$\mathbb{E}^*_f(X) = x_0 \qquad \text{and} \qquad \mathbb{V}^*_f(X) = \frac{\gamma^2}{2},$$

with location parameter $x_0 \in \mathbb{R}$ and scale parameter $\gamma > 0$

Central limit theorem

Variance

Variance of an u.v.

CLT suggests a definition of the variance as

$$\mathbb{V}^*_f(X) = -\frac{1}{f''(\mathbb{E}^*_f(X))}$$

Example (Gamma possibility function on $(0, \infty)$)

$$\mathbb{E}^*_f(X) = k\theta \qquad \text{and} \qquad \mathbb{V}^*_f(X) = k\theta^2,$$

with shape parameter $k \geq 0$ and scale parameter $\theta > 0$

Example (Cauchy possibility function on $\mathbb{R}$)

$$\mathbb{E}^*_f(X) = x_0 \qquad \text{and} \qquad \mathbb{V}^*_f(X) = \frac{\gamma^2}{2},$$

with location parameter $x_0 \in \mathbb{R}$ and scale parameter $\gamma > 0$

Outline

Bayesian inference

Ingredients:

- Unknown: u.v. X on $\mathbf{X}$ represented by a prior f
- Observation: u.v. Y represented by a likelihood $h(\cdot \,|\, x)$

Bayes' rule

$$f(x \,|\, y) = \frac{h(y \,|\, x) f(x)}{\sup_{x' \in \mathbf{X}} h(y \,|\, x') f(x')}$$

for some realisation y of Y

Connection with optimisation

Estimating $\mathbb{E}^*_{f(\cdot \,|\, y)}(X)$ is the same as

$$x^* = \operatorname*{argmax}_{x \in \mathbf{X}} \log h(y \,|\, x) + \log f(x)$$

with $\log f$ some regularisation

Bayesian inference

Ingredients:

- Unknown: u.v. X on $\mathbf{X}$ represented by a prior f
- Observation: u.v. Y represented by a likelihood $h(\cdot \,|\, x)$

Bayes' rule

$$f(x \,|\, y) = \frac{h(y \,|\, x) f(x)}{\sup_{x' \in \mathbf{X}} h(y \,|\, x') f(x')}$$

for some realisation y of Y

Connection with optimisation

Estimating $\mathbb{E}^*_{f(\cdot \,|\, y)}(X)$ is the same as

$$x^* = \operatorname*{argmax}_{x \in \mathbf{X}} \log h(y \,|\, x) + \log f(x)$$

with $\log f$ some regularisation

Bayesian inference

Ingredients:

- Unknown: u.v. X on $\mathbf{X}$ represented by a prior f
- Observation: u.v. Y represented by a likelihood $h(\cdot \,|\, x)$

Bayes' rule

$$f(x \,|\, y) = \frac{h(y \,|\, x) f(x)}{\sup_{x' \in \mathbf{X}} h(y \,|\, x') f(x')}$$

for some realisation y of Y

Connection with optimisation

Estimating $\mathbb{E}^*_{f(\cdot \,|\, y)}(X)$ is the same as

$$x^* = \operatorname*{argmax}_{x \in \mathbf{X}} \log h(y \,|\, x) + \log f(x)$$

with $\log f$ some regularisation

Bayesian inference

Ingredients:

- Unknown: u.v. X on $\mathbf{X}$ represented by a prior f
- Observation: u.v. Y represented by a likelihood $h(\cdot \,|\, x)$

Bayes' rule

$$f(x \,|\, y) = \frac{h(y \,|\, x) f(x)}{\sup_{x' \in \mathbf{X}} h(y \,|\, x') f(x')}$$

for some realisation y of Y

Connection with optimisation

Estimating $\mathbb{E}^*_{f(\cdot \,|\, y)}(X)$ is the same as

$$x^* = \operatorname*{argmax}_{x \in \mathbf{X}} \log h(y \,|\, x) + \log f(x)$$

with $\log f$ some regularisation

Filtering for possibility functions

A state space model

Consider a partially-observed Markov chain $\{X_n\}_n$ on $\mathbf{X}$ such that

$$X_n = G(X_{n-1}) + V_n$$
$$Y_n = H(X_n) + W_n$$

with $\{V_n\}_n$ and $\{W_n\}_n$ i.i.d. such that

$$f_{X_n}(\cdot \mid X_{n-1}) = g(\cdot \mid X_{n-1}) \qquad \text{and} \qquad f_{Y_n}(\cdot \mid X_n) = h(\cdot \mid X_n)$$

Filtering equations

$$\begin{cases} f_{X_n}(x \mid y_{1:n-1}) = \sup_{x' \in \mathbf{X}} g(x \mid x') f_{X_{n-1}}(x' \mid y_{1:n-1}) \\ f_{X_n}(x \mid y_{1:n}) = \dfrac{h(y_n \mid x) f_{X_n}(x \mid y_{1:n-1})}{\sup_{x' \in \mathbf{X}} h(y_n \mid x') f_{X_n}(x' \mid y_{1:n-1})}. \end{cases}$$

Filtering for possibility functions

A state space model

Consider a partially-observed Markov chain $\{X_n\}_n$ on $\mathbf{X}$ such that

$$X_n = G(X_{n-1}) + V_n$$
$$Y_n = H(X_n) + W_n$$

with $\{V_n\}_n$ and $\{W_n\}_n$ i.i.d. such that

$$f_{X_n}(\cdot \,|\, X_{n-1}) = g(\cdot \,|\, X_{n-1}) \qquad \text{and} \qquad f_{Y_n}(\cdot \,|\, X_n) = h(\cdot \,|\, X_n)$$

Filtering equations

$$\begin{cases} f_{X_n}(x \,|\, y_{1:n-1}) = \sup\limits_{x' \in \mathbf{X}} g(x \,|\, x') f_{X_{n-1}}(x' \,|\, y_{1:n-1}) \\[2ex] f_{X_n}(x \,|\, y_{1:n}) = \dfrac{h(y_n \,|\, x) f_{X_n}(x \,|\, y_{1:n-1})}{\sup_{x' \in \mathbf{X}} h(y_n \,|\, x') f_{X_n}(x' \,|\, y_{1:n-1})}. \end{cases}$$

Kalman filter

Recursion

$$f_{n-1}(x \mid y_{1:n-1}) = \bar{\mathcal{N}}(x; m_{n-1}, \Sigma_{n-1})$$
$$g(x \mid x') = \bar{\mathcal{N}}(x; Fx', Q)$$
$$h(y \mid x) = \bar{\mathcal{N}}(y; Hx, R)$$

$\rightarrow$ Same means $m_{n|n-1}$, m_n and spreads $\Sigma_{n|n-1}$, Σ_n

Different marginal likelihood

$$f_{Y_n}(y_n) = \exp\Big(-\frac{1}{2}(y_n - Hm_{n|n-1})^T S_n^{-1}(y_n - Hm_{n|n-1})\Big)$$

with $S_n = H\Sigma_{n|n-1}H^T + R$

J. H. and A. Bishop. "Smoothing and filtering with a class of outer measures". In: *SIAM Journal on Uncertainty Quantification* 6.2 (2018)

Kalman filter

Recursion

$$f_{n-1}(x \mid y_{1:n-1}) = \bar{\mathcal{N}}(x; m_{n-1}, \Sigma_{n-1})$$
$$g(x \mid x') = \bar{\mathcal{N}}(x; Fx', Q)$$
$$h(y \mid x) = \bar{\mathcal{N}}(y; Hx, R)$$

$\rightarrow$ Same means $m_{n|n-1}$, m_n and spreads $\Sigma_{n|n-1}$, Σ_n

Different marginal likelihood

$$f_{Y_n}(y_n) = \exp\Big(-\frac{1}{2}(y_n - Hm_{n|n-1})^T S_n^{-1}(y_n - Hm_{n|n-1})\Big)$$

with $S_n = H\Sigma_{n|n-1}H^T + R$

J. H. and A. Bishop. "Smoothing and filtering with a class of outer measures". In: *SIAM Journal on Uncertainty Quantification* 6.2 (2018)

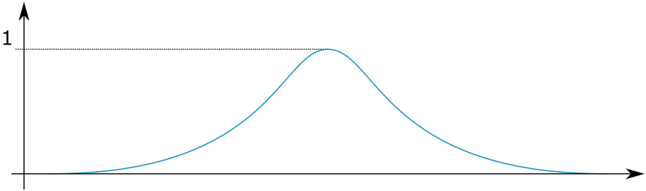
Non-linear filtering

Particle approximation of f:

1. Find the probability distribution

$$p^* = \operatorname*{argmax}_{p \in \mathcal{P}(\mathbf{X})} \mathrm{H}(p)$$

   such that $\mathbf{1}_B(x)p(x)\mathrm{d}x \leq \sup_{x \in B} f(x)$ for any $B \subseteq \mathbf{X}$
2. Get N samples $x_1, \ldots, x_n$ from p^*
3. Weight the samples as $w_i = f(x_i)$, $i \in \{1, \ldots, N\}$



J. H. and B. Ristić. "Sequential Monte Carlo algorithms for a class of outer measures". In: *arXiv preprint arXiv:1704.01233* (2017)

Non-linear filtering

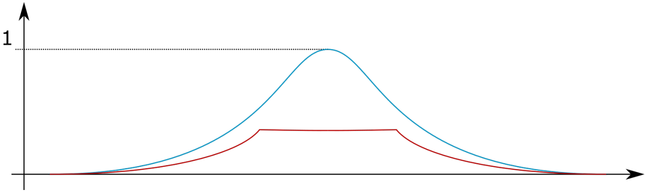
Particle approximation of f:

1. Find the probability distribution

$$p^* = \operatorname*{argmax}_{p \in \mathcal{P}(\mathbf{X})} \mathrm{H}(p)$$

   such that $\mathbf{1}_B(x)p(x)\mathrm{d}x \leq \sup_{x \in B} f(x)$ for any $B \subseteq \mathbf{X}$

2. Get N samples $x_1, \ldots, x_n$ from p^*
3. Weight the samples as $w_i = f(x_i)$, $i \in \{1, \ldots, N\}$



J. H. and B. Ristić. "Sequential Monte Carlo algorithms for a class of outer measures". In: *arXiv preprint arXiv:1704.01233* (2017)

Non-linear filtering

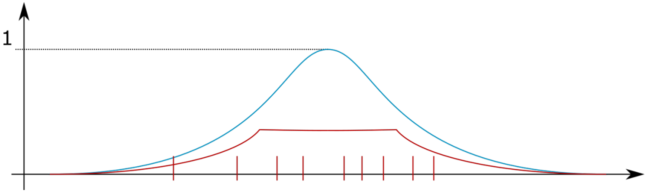
Particle approximation of f:

1. Find the probability distribution

$$p^* = \operatorname*{argmax}_{p \in \mathcal{P}(\mathbf{X})} \mathrm{H}(p)$$

such that $\mathbf{1}_B(x)p(x)\mathrm{d}x \leq \sup_{x \in B} f(x)$ for any $B \subseteq \mathbf{X}$

2. Get N samples $x_1, \ldots, x_n$ from p^*
3. Weight the samples as $w_i = f(x_i)$, $i \in \{1, \ldots, N\}$



J. H. and B. Ristić. "Sequential Monte Carlo algorithms for a class of outer measures". In: *arXiv preprint arXiv:1704.01233* (2017)

Non-linear filtering

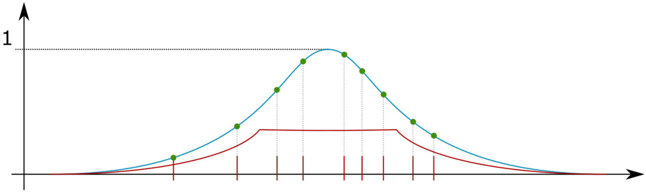
Particle approximation of f:

1. Find the probability distribution

$$p^* = \operatorname*{argmax}_{p \in \mathcal{P}(\mathbf{X})} \mathrm{H}(p)$$

   such that $\mathbf{1}_B(x) p(x) \mathrm{d}x \leq \sup_{x \in B} f(x)$ for any $B \subseteq \mathbf{X}$

2. Get N samples $x_1, \ldots, x_n$ from p^*
3. Weight the samples as $w_i = f(x_i)$, $i \in \{1, \ldots, N\}$



J. H. and B. Ristić. "Sequential Monte Carlo algorithms for a class of outer measures". In: *arXiv preprint arXiv:1704.01233* (2017)

Bearings-only tracking: Robustness

Standard scenario with measurement noise:

- modelled as Gaussian
- simulated as Student's t with ν d.f.

N = 2000

ν	3	5	8	∞
probability	43.6	21.2	12.4	5.4
possibility	21.0	6.2	1.8	0.6

N = 5000

ν	3	5	8	∞
probability	34.0	11.6	6.2	0.6
possibility	17.8	2.8	0.4	0

B. Ristić, J. H., and S. Arulampalam. "Robust target motion analysis using the possibility Particle filter". In: *IET Radar Sonar Navigation* (2018)

Bearings-only tracking: Robustness

Standard scenario with measurement noise:

- modelled as Gaussian
- simulated as Student's t with ν d.f.

N = 2000

ν	3	5	8	∞
probability	43.6	21.2	12.4	5.4
possibility	21.0	6.2	1.8	0.6

N = 5000

ν	3	5	8	∞
probability	34.0	11.6	6.2	0.6
possibility	17.8	2.8	0.4	0

B. Ristić, J. H., and S. Arulampalam. "Robust target motion analysis using the possibility Particle filter". In: *IET Radar Sonar Navigation* (2018)

Bearings-only tracking: Robustness

Standard scenario with measurement noise:

- modelled as Gaussian
- simulated as Student's t with ν d.f.

N = 2000

ν	3	5	8	∞
probability	43.6	21.2	12.4	5.4
possibility	21.0	6.2	1.8	0.6

N = 5000

ν	3	5	8	∞
probability	34.0	11.6	6.2	0.6
possibility	17.8	2.8	0.4	0

B. Ristić, J. H., and S. Arulampalam. "Robust target motion analysis using the possibility Particle filter". In: *IET Radar Sonar Navigation* (2018)

Natural language processing: Bike theft

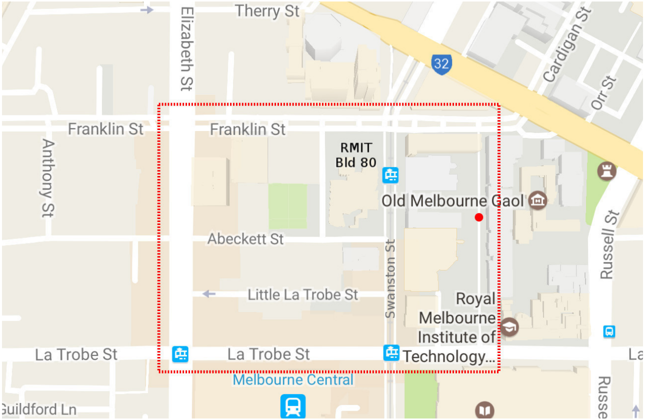


Figure: Map of the surroundings (Google Maps). Red-dotted rectangle: area of interest, red dot: location of bike theft.

A. Bishop, J. H., D. Angley, and B. Ristić. "Spatio-temporal tracking from natural language statements using outer probability theory". In: *Elsevier Information Sciences* 463–464 (2018)

Natural language processing: Bike theft

Information to be confirmed:

1. *Suspect alibi*: "I was with a friend at the tram stop on the intersection of La Trobe St. and Elizabeth St."
2. *CCTV*: Recording of the theft

The witnesses' declarations are:

1. "The suspect has been seen on Elizabeth St. around 2.07p.m."
2. "The suspect turned at the intersection of Swanston and Abeckett St. between 2.25p.m. and 2.35p.m."
3. "The suspect has been seen near RMIT building 80 around 2.35p.m."

Natural language processing: Bike theft

Information to be confirmed:

1. *Suspect alibi*: “I was with a friend at the tram stop on the intersection of La Trobe St. and Elizabeth St.”
2. *CCTV*: Recording of the theft

The witnesses’ declarations are:

1. “The suspect has been seen on Elizabeth St. around 2.07p.m.”
2. “The suspect turned at the intersection of Swanston and Abeckett St. between 2.25p.m. and 2.35p.m.”
3. “The suspect has been seen near RMIT building 80 around 2.35p.m.”


Franklin St
Old Melbourne
Abeckett St
Little La Trobe St
Royal
Melbourne
Institute of
La Trobe St
Technology...

A negative example: track before detect

Principle:

- Use full image instead of extracted point measurements
- Model the background noise and the signal directly

Attempt:

- Use possibility function to bring robustness
- Not informative enough

Possible direction:

→ Use outer probability measures of the form

$$\bar{P}(B) = \sup_{\theta \in B} p(B \mid \theta) f(\theta)$$

→ More informative model that still retains some robustness

A negative example: track before detect

Principle:

- Use full image instead of extracted point measurements
- Model the background noise and the signal directly

Attempt:

- Use possibility function to bring robustness
- Not informative enough

Possible direction:

→ Use outer probability measures of the form

$$\bar{P}(B) = \sup_{\theta \in B} p(B \,|\, \theta) f(\theta)$$

→ More informative model that still retains some robustness

A negative example: track before detect

Principle:

- Use full image instead of extracted point measurements
- Model the background noise and the signal directly

Attempt:

- Use possibility function to bring robustness
- Not informative enough

Possible direction:

→ Use outer probability measures of the form

$$\bar{P}(B) = \sup_{\theta \in B} p(B \mid \theta) f(\theta)$$

→ More informative model that still retains some robustness

Outline

Back to the motivating example

Multi-target tracking with possibility functions

Ingredients:

- Uncertain counting measure described by $\boldsymbol{f}_n$
- Represented by the intensity measure

$$\bar{F}_n(B) = \bar{\mathbb{E}}_{\boldsymbol{f}_n}\Big(\max_{1 \leq i \leq N} \mathbf{1}_B(X_n^i) \Big)$$

Update equation:

$$\bar{F}_n(x \,|\, \mathcal{Y}_{1:n}) = a_{\mathrm{FN}} \bar{F}_n(x \,|\, \mathcal{Y}_{1:n-1}) \vee \max_{y \in \mathcal{Y}_n} \frac{a_{\mathrm{TP}} h(y \,|\, x) \bar{F}_n(x \,|\, \mathcal{Y}_{1:n-1})}{\bar{F}_{\mathrm{FP}}(y) \vee a_{\mathrm{TP}} \sup_{x'} h(y \,|\, x') \bar{F}_n(x' \,|\, \mathcal{Y}_{1:n-1})}$$

with $a_{\mathrm{FN}} \vee a_{\mathrm{TP}} = 1$

J. H. "Detection and estimation of partially-observed dynamical systems: an outer-measure approach". In: *arXiv:1801:00571* (2018)

Back to the motivating example

Multi-target tracking with possibility functions

Ingredients:

- Uncertain counting measure described by $\boldsymbol{f}_n$
- Represented by the intensity measure

$$\bar{F}_n(B) = \bar{\mathbb{E}}_{\boldsymbol{f}_n}\Big(\max_{1 \leq i \leq N} \mathbf{1}_B(X_n^i) \Big)$$

Update equation:

$$\bar{F}_n(x \,|\, \mathcal{Y}_{1:n}) = a_{\mathrm{FN}} \bar{F}_n(x \,|\, \mathcal{Y}_{1:n-1}) \vee \max_{y \in \mathcal{Y}_n} \frac{a_{\mathrm{TP}} h(y \,|\, x) \bar{F}_n(x \,|\, \mathcal{Y}_{1:n-1})}{\bar{F}_{\mathrm{FP}}(y) \vee a_{\mathrm{TP}} \sup_{x'} h(y \,|\, x') \bar{F}_n(x' \,|\, \mathcal{Y}_{1:n-1})}$$

with $a_{\mathrm{FN}} \vee a_{\mathrm{TP}} = 1$

J. H. “Detection and estimation of partially-observed dynamical systems: an outer-measure approach”. In: *arXiv:1801:00571* (2018)

Back to the motivating example

Multi-target tracking with possibility functions

Advantages:

- Only a lower bound on the probability of detection is required, e.g.

$$a_{\mathrm{FN}} = 0.2 \text{ and } a_{\mathrm{TP}} = 1 \implies p_{\mathrm{TP}} \in [0.8, 1]$$

- The distribution of the false positives does not need to be fully known

Challenges:

- Track extraction is less obvious
- At least *some* information is required on the false positives

$\rightarrow$ proceed by negation?

J. H. "Detection and estimation of partially-observed dynamical systems: an outer-measure approach". In: *arXiv:1801:00571* (2018)

Back to the motivating example

Multi-target tracking with possibility functions

Advantages:

- Only a lower bound on the probability of detection is required, e.g.

$$a_{\mathrm{FN}} = 0.2 \text{ and } a_{\mathrm{TP}} = 1 \implies p_{\mathrm{TP}} \in [0.8, 1]$$

- The distribution of the false positives does not need to be fully known

Challenges:

- Track extraction is less obvious
- At least *some* information is required on the false positives

$\rightarrow$ proceed by negation?

J. H. "Detection and estimation of partially-observed dynamical systems: an outer-measure approach". In: *arXiv:1801:00571* (2018)

Concluding remarks

- Most of the introduced ideas extend to outer probability measures like $\bar{P}$
- Discussions about the nature of uncertainty become practical
- There is an opportunity to provide justification for heuristics

Concluding remarks

- Most of the introduced ideas extend to outer probability measures like $\bar{P}$
- Discussions about the nature of uncertainty become practical
- There is an opportunity to provide justification for heuristics

Concluding remarks

- Most of the introduced ideas extend to outer probability measures like $\bar{P}$
- Discussions about the nature of uncertainty become practical
- There is an opportunity to provide justification for heuristics

References

A. Bishop, J. H., D. Angley, and B. Ristić. "Spatio-temporal tracking from natural language statements using outer probability theory". In: *Elsevier Information Sciences* 463–464 (2018).

J. H. "Detection and estimation of partially-observed dynamical systems: an outer-measure approach". In: *arXiv:1801:00571* (2018).

J. H. "Parameter estimation with a class of outer probability measures". In: *arXiv:1801.00569* (2018).

J. H. and A. Bishop. "Smoothing and filtering with a class of outer measures". In: *SIAM Journal on Uncertainty Quantification* 6.2 (2018).

J. H. and B. Ristić. "Sequential Monte Carlo algorithms for a class of outer measures". In: *arXiv preprint arXiv:1704.01233* (2017).

R.P.S. Mahler. "Multitarget Bayes filtering via first-order multitarget moments". In: *IEEE Transactions on Aerospace and Electronic systems* 39.4 (2003).

B. Ristić, J. H., and S. Arulampalam. "Robust target motion analysis using the possibility Particle filter". In: *IET Radar Sonar Navigation* (2018).

Thank you!

jeremie.houssineau@warwick.ac.uk
`jeremiehoussineau.com`