# Constructing consonant beliefs from multivariate data with scenario theory

Marco De Angelis[1], Roberto Rocchetta[2], Ander Gray[1], Scott Ferson[1]

[1] Institute for Risk and Uncertainty, University of Liverpool
[2] Eindhoven University of Technology

ISIPTA 2021, virtually in Granada

July 7, 2021

# Outline

# Problem statement

Given a bunch of *iid* samples $X_1, ..., X_n$, with $X_i \in \mathbb{R}^m$, what can be learnt about the unknown underlying distribution $\mathbb{P}_X$? What is the probability of observing a new sample in a given set?

# Problem statement

Given a bunch of *iid* samples $X_1, ..., X_n$, with $X_i \in \mathbb{R}^m$, what can be learnt about the unknown underlying distribution $\mathbb{P}_X$? What is the probability of observing a new sample in a given set?

- Samples are few

# Problem statement

Given a bunch of *iid* samples $X_1, ..., X_n$, with $X_i \in \mathbb{R}^m$, what can be learnt about the unknown underlying distribution $\mathbb{P}_X$? What is the probability of observing a new sample in a given set?

- Samples are few
- Multivariate case

# Problem statement

Given a bunch of *iid* samples $X_1, ..., X_n$, with $X_i \in \mathbb{R}^m$, what can be learnt about the unknown underlying distribution $\mathbb{P}_X$? What is the probability of observing a new sample in a given set?

- Samples are few — We cannot learn $\mathbb{P}_X$ exactly!
- Multivariate case

# Problem statement

Given a bunch of *iid* samples $X_1, ..., X_n$, with $X_i \in \mathbb{R}^m$, what can be learnt about the unknown underlying distribution $\mathbb{P}_X$? What is the probability of observing a new sample in a given set?

- Samples are few — We cannot learn $\mathbb{P}_X$ exactly!
- Multivariate case — What about $X_i$ interdependence?

# Predictive beliefs

We want to obtain a belief function $\text{Bel}_X$ so that the inequality $\text{Bel}_X \leq \mathbb{P}_X$ holds at least $100\ (1-\beta)\%$ of the times.

# Predictive beliefs

We want to obtain a belief function $\text{Bel}_X$ so that the inequality $\text{Bel}_X \leq \mathbb{P}_X$ holds at least $100\ (1-\beta)\%$ of the times.

1. $\forall A \subseteq \mathbb{R}^m,\ \text{Bel}_X(A) \rightarrow \mathbb{P}_X(A),\ n \rightarrow \infty$

# Predictive beliefs

We want to obtain a belief function $\mathsf{Bel}_X$ so that the inequality $\mathsf{Bel}_X \leq \mathbb{P}_X$ holds at least $100\,(1-\beta)\%$ of the times.

1. $\forall A \subseteq \mathbb{R}^m,\ \mathsf{Bel}_X(A) \to \mathbb{P}_X(A),\ n \to \infty$
2. $\mathbb{P}^n\left(\mathsf{Bel}_X \leq \mathbb{P}_X\right) \geq 1-\beta$

# Coherent lower probabilities

Basic mass assignments: (i) $m(\emptyset) = 0$, (ii) $\sum_{A \in 2^{\mathcal{X}}} m(A) = 1$

Beliefs obtained from *basic mass assignments* are coherent lower probabilities.

The subsets $A \subseteq \mathcal{X}$ such that $m(A) > 0$ are called *focal elements*. The belief of a focal set $A$, for all $B \in 2^{\mathcal{X}}$, is

$$\text{Bel}_X(A) = \sum_{B: B \subseteq A} m(B). \quad (1)$$

# Scenario optimization

Let $z \in \mathcal{Z} \subseteq \mathbb{R}^d$ be a vector of (design) parameters and $X_1, ..., X_n$ a bunch of *iid* samples , with $X_i \in R^m$. The scenario optimization consists in minimizing the convex cost function $f : \mathcal{Z} \to \mathbb{R}$:

$$\min_{z \in \mathcal{Z}} f(z)$$
$$\text{subject to: } z \in \bigcap_{i=1,...,n} \mathcal{Z}_{X_i}, \qquad (2)$$

# Scenario optimization

Let $z \in \mathcal{Z} \subseteq \mathbb{R}^d$ be a vector of (design) parameters and $X_1, ..., X_n$ a bunch of *iid* samples , with $X_i \in R^m$. The scenario optimization consists in minimizing the convex cost function $f : \mathcal{Z} \rightarrow \mathbb{R}$:

$$\min_{z \in \mathcal{Z}} f(z)$$
$$\text{subject to: } z \in \bigcap_{i=1,...,n} \mathcal{Z}_{X_i}, \tag{2}$$

Design parameters can be the center coordinates and the radius of a circle ($\mathbb{R}^{m=2}$) or sphere ($\mathbb{R}^{m=3}$), as it will be illustrated in the next slide.

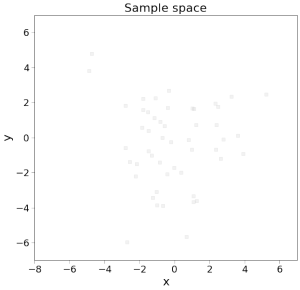
# Scenario optimization on the disk $\mathbb{R}^2$

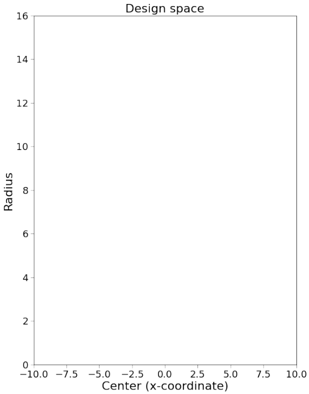
For example, let $z = (c_x, r)$, be the centre x-coordinate and the radius of a circle ($c_y = 0$). The scenario optimization consists in minimizing the area of the circle:

$$\begin{aligned} \min_{(c_x,r)} \quad & \pi r^2 \\ \text{subject to:} \quad & (c_x - X_1)^2 + (0 - Y_1)^2 \leq r^2, \\ & \dots \\ & (c_x - X_n)^2 + (0 - Y_n)^2 \leq r^2 \end{aligned} \tag{3}$$

# Visualizing the constraints: $m = 2, d = 2$

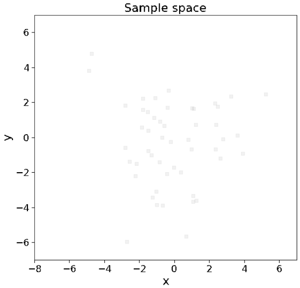
Center y-coordinate = $0$

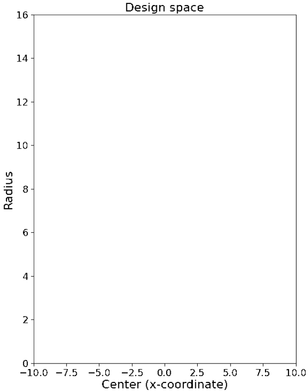


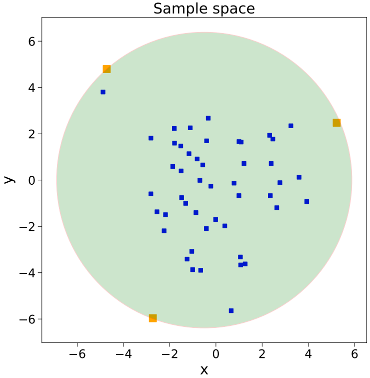
# Obtaining the smallest disk $O(n)$

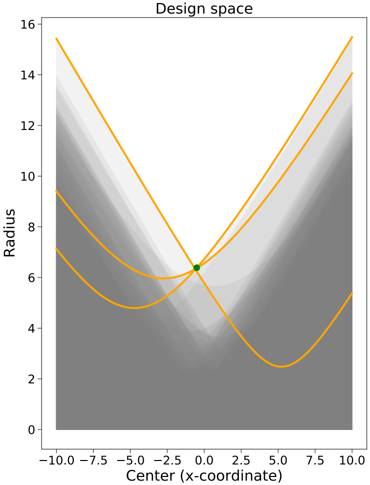
Center y-coordinate = $0$



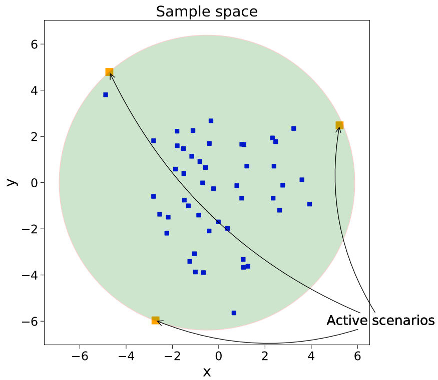



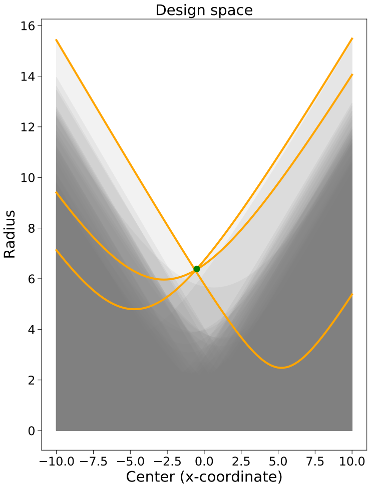
# Active scenarios



# Active scenarios



# Definitions

**Enclosing set of degree $k$**: The optimal set $B_k \subseteq \mathbb{R}^m$, that strictly contains $n-k$ observations.

**Lower probability of enclosing set $B_k$**: The precise predictive probability of a given enclosing set of degree $k$, $\mathbb{P}_X(B_k)$, has a lower bound $\underline{p}_k$, with assigned one-sided coverage probability.

$$\mathbb{P}^n\left(\underline{p}_k \leq \mathbb{P}_X(B_k)\right) \geq 1-\beta, \quad (4)$$

# Computing the lower bound

$$\varphi(t) = \frac{\beta}{n+1}\sum_{j=k}^{n}\binom{j}{k}t^{j-k} - \binom{n}{k}t^{n-k}, \quad t \in [0,1] \tag{5}$$

$$\varphi(\hat{t}) = 0; \qquad \underline{p}_k = \hat{t}(n,k,\beta);$$

$$\mathbb{P}^n\left(\underline{p}_k \leq \mathbb{P}_X(B_k)\right) \geq 1-\beta$$

Campi, M.C. and Garatti, S., 2018. *Wait-and-judge scenario optimization*. Mathematical Programming, 167(1), pp.155-189.

Garatti, S. and Campi, M.C., 2019. *Risk and complexity in scenario optimization*. Mathematical Programming, pp.1-37.

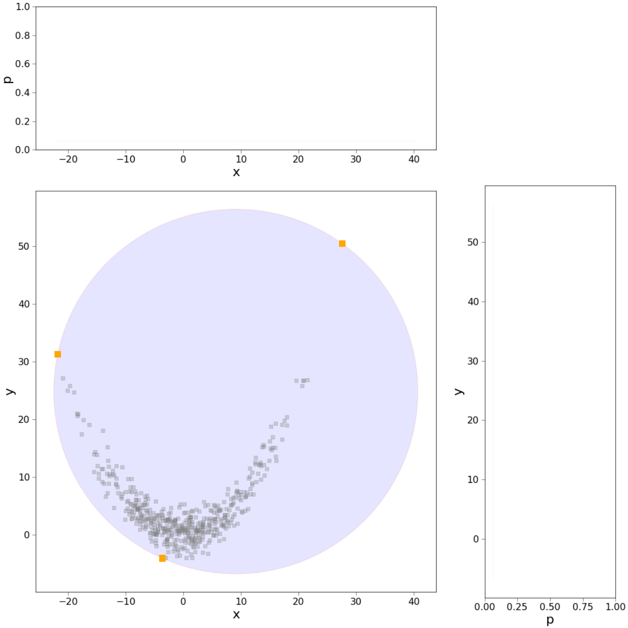
**Theorem 1**

*The lower bounds $\underline{p}_k$ make a sequence of coherent predictive beliefs for any $k \in \mathbb{Z}_+$ such that $0 = k_0 < k_1 < \cdots < k_n = n$.*

**Theorem 1**

*The lower bounds $\underline{p}_k$ make a sequence of coherent predictive beliefs for any $k \in \mathbb{Z}_+$ such that $0 = k_0 < k_1 < \cdots < k_n = n$.*

The proof follows from Eq.(5), noticing that the roots of the polynomial are decreasing with $k$.


p
x
y
p
y
x

# Conclusions

- Inference on multidimensional datasets
- No need to estimate the likelihood
- Additional constraints can ensure sets are fully inter-nested
- Structures can be propagated and retain the confidence interpretation
- The interdependence is encoded in the shape of the enclosing sets