Constrained probabilistic modeling with OpenTURNS

R. Lebrun (Airbus)

User Day #17, June 14th 2024, EDF Lab

Overview

Taking constraints into account:

- Order statistics
- Distributions over meshes
- Meshing capabilities

Order statistics I

Let F_X be the CDF of a continuous random variable X and $(X_1, \dots, X_n)$ a sample of size n of X.
The order statistics associated to this sample is the (almost surely) unique increasing reordering $(X_{(1)}, \dots, X_{(n)})$ of this sample: there exists a permutation $\sigma \in \mathrm{t}_n$ such that $X_{(k)} = X_{\sigma(k)}$ and

$$X_{(1)} \leq \cdots \leq X_{(n)}$$

It can be shown that the joint CDF of $(X_{(1)}, \dots, X_{(n)})$ writes:

$$F_{(X)}(x_1, \dots, x_n) = F_{(U)}(F_X(x_1), \dots, F_X(x_n))$$

where $F_{(U)}$ is the joint CDF of the order statistics associated to a sample of size n of $U \simeq \mathcal{U}(0,1)$.
This relation is exactly the relation between a joint CDF, its copula and its marginal distributions.

Order statistics II

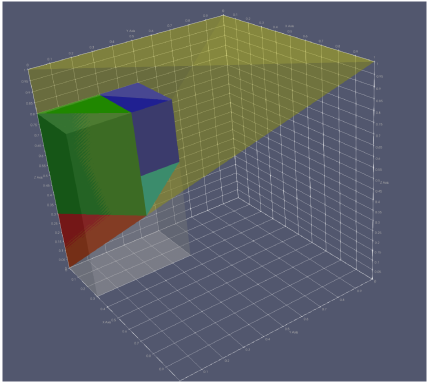
$\Longrightarrow$ the ComposedDistribution class has been renamed to JointDistribution and has been extended to define a multivariate CDF F given n univariate CDFs $F_1, \dots, F_n$ and a core K, which is a multivariate CDF of dimension n which range is included into $[0,1]^n$:

$$F(x_1, \dots, x_n) = K(F_1(x_1), \dots, F_n(x_n))$$

If K is a copula, then $F_1, \dots, F_n$ are the marginal distributions of F.
The specific choice $K = F_{(U)}$ and $F_1 = \cdots = F_n = F_X$ leads to the joint distribution of the order n statistics associated to F_X.
The new UniformOrderStatistics class corresponds to $F_{(U)}$.

Order statistics III

The computation of $F_{(U)}$ is tricky. The associated support is the simplex $\mathcal{S} = \{(u_1, \dots, u_n) \in [0,1]^n \,|\, 0 \leq u_1 \leq \cdots \leq u_n\}$ and the computation resorts to compute the volume of the intersction between $\mathcal{S}$ and $[0, u_1] \times \cdots \times [0, u_n]$, which can be done recursively.



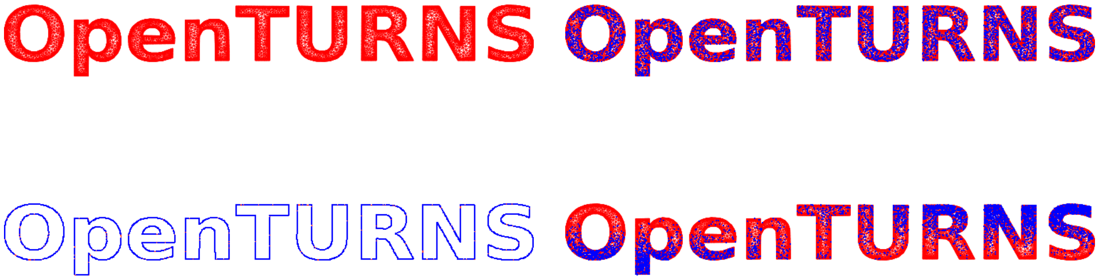
Distributions over a mesh I

Let $\mathcal{D} \subset \mathbb{R}^d$ be a complex domain given as a `Mesh`.

- How to sample it uniformly? See the `UniformOverMesh` distribution.
  - Sampling cost constant wrt the number of simplices
  - Sampling cost proportional to the dimension
  - PDF computed using an efficient kd-tree $\rightarrow \mathcal{O}(n_{vertices})$
  - CDF computed using numerical integration $\rightarrow$ limited to low dimension

  can be used in high dimension settings.
- How to sample it nonuniformly? See the `TruncatedOverMesh` distribution.
  - Sampling done using rejection $\rightarrow$ limited to low dimension
  - PDF computed using an efficient kd-tree $\rightarrow \mathcal{O}(n_{vertices})$ then a numerical integration using `SimplicialCubature` limited to low dimension
  - CDF computed using numerical integration over the PDF computation $\rightarrow$ very expansive

Distributions over a mesh II



New meshing capabilities I

- The `IntervalMesher` class has been extended to any dimension
- The `LevelSetMesher` class has been extended to any dimension
- A new projection algorithm has been added to the `LevelSetMesher` class to get a an accuracy of the order of the machine precision (mandatory for some applications)
- The `BoundaryMesher` class allows to extract the boundary mesh of dimension $d-1$ of any mesh of dimension d. It also proposes a "thick" version, where the faces are replaced by simplices of given height.

New meshing capabilities II