OpenTURNS release highlights

J. Schueller (Phimeca)

User Day #16, June 23rd 2023, EDF Lab

New features since last year in releases:

- v1.20: fall 2022
- v1.21: ~~spring~~ summer 2023

Smolyak quadratures

- Combination of tensorized quadratures to build an n-d DOE from n 1-d DOEs
- The combination technique reduces the nodes number wrt raw tensorization rule
- Fine-grained control over the nodes/weights merging step, access m-indices
- Designed to be used for functional chaos (by integration or least-squares)
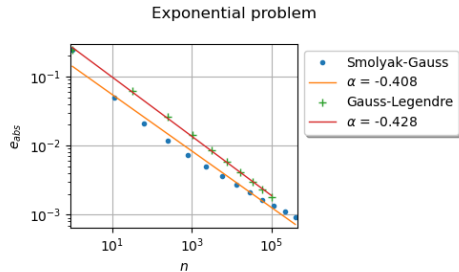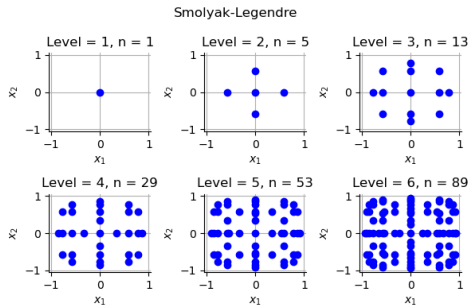
Cross-entropy importance sampling 1/2

- Adaptive, importance-sampling based simulation algorithm
- Update an intermediate threshold q_k at each step

$$\rho \in [0,1], q_k = \max(T, y^{(k)}_{\lfloor \rho N \rfloor})$$

- Auxiliary distribution h parameters $\underline{\lambda}$ numerically optimized at each step

$$\underline{\lambda}_k = \mathrm{argmax}_\lambda \frac{1}{N} \sum_{i=1}^{N} \mathbb{1}_{g(\underline{x}_i^{(k)}) \leq q_k} \frac{f_{\underline{X}}(\underline{x}_i^{(k)})}{h_{\underline{\lambda}_{k-1}}(\underline{x}_i^{(k)})} \log(h_{\underline{\lambda}}(\underline{x}_i^{(k)}))$$

- Compute the final importance probability

$$\widehat{P}^{CE}(g(\underline{\underline{X}}) < T) = \frac{1}{N} \sum_{i=1}^{N} \mathbb{1}_{g(\underline{x}_i^{(k)}) < T} \frac{f_{\underline{X}}(\underline{x}_i^{(k)})}{h_{\underline{\lambda}_{k-1}}(\underline{x}_i^{(k)})}$$

Cross-entropy importance sampling 2/2

- Standard space variant with analytical optimum of the Normal parameters
- Very good performance, can be compared to subset
- Also, no COV-based termination criterion (based on threshold)

Cloud of samples and failure domain
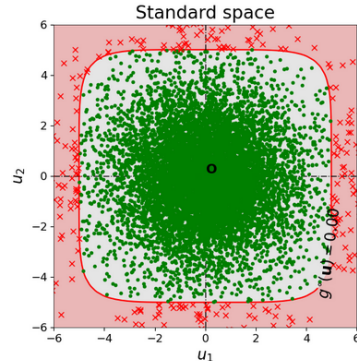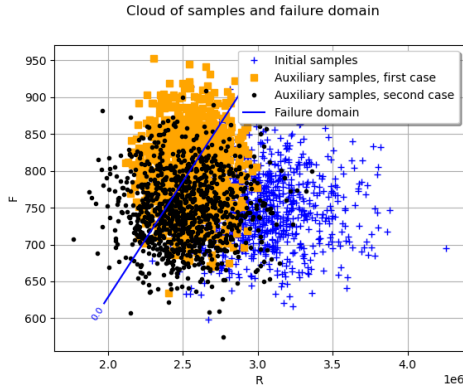
Uniform distribution over a mesh

- Uniform sampling in any closed set delimited by a generic mesh, any dimension
- Sampling of simplices weighted by volume, then uniform sampling inside simplex
- CDF is obtained by integration at mesh bounds
- Efficient implementation, low-level primitives for tetras, etc ...
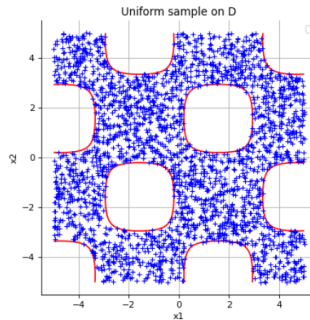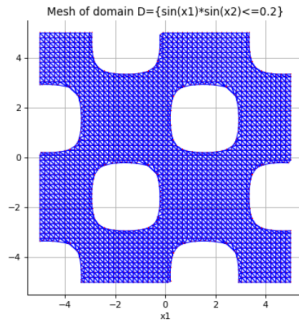
HSIC performance improvements

- Loop reordering and parallelization of P-Values evaluation (permutation) using TBB
- Efficient trace computation in HSICVStat instead of full products
- Cache input covariance matrix discretization
- Use STL primitives like std::accumulate

Benchmark (global HSIC, 3 OT versions, 2 estimators, 2 sizes, 2 p-values types)

| version/ustat|vstat	1.19/U	1.20/U	1.21/U	1.19/V	1.20/V	1.21/V
time (100/perm)	1.30	1.31	0.082	6.65	6.70	0.01
time (1000/perm)	-	762.8	36.58	-	5313.4	0.79
time (100/asymp)	0.028	0.028	0.0013	0.028	0.031	0.0017
time (1000/asymp)	16.59	16.46	0.17	16.56	17.85	0.15

Chaos rewrite

- Long-term effort to simplify usage and improve performance
- Deduplicate basis selection at AdaptiveStrategy and LARS level
- Restrict combination of integration + LARS
- New LeastSquaresExpansion and IntegrationExpansion classes available in 1.21
- Ongoing work to rewrite LARS selection as well (2024 ?)

Selection history
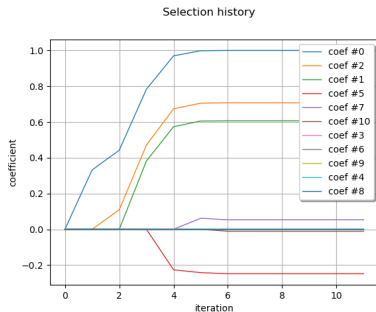
Documentation improvements

- Lots of new examples: chaos, cv, regression, MLE, functions, integration, enumerate, ...
- New usecases: fire satellite, wing weight, Linthurst/Coles datasets
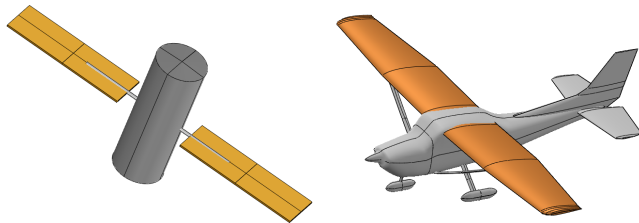
- Example minigalleries linking to relevant examples
- Lot of time invested in the improvement of the documentation

Other improvements

- Improved MCMC adaptation and new class to customize the update
- Field to vector metamodeling and sensitivity using KL + chaos
- Chaos expansion for mixed variables
- New BoxCoxFactory method to return directly the underlying linear model
- New inference method based on quantile numerical optimization
- Enabled Pagmo.moead_gen, Bonmin.Ecp/iFP optimization algorithms
- Continued bugfix effort (avg 30+ recorded per release + more internally fixed)

Python channels

- Pip, Conda
- Versions: 3.8-3.11
- OS: Windows, Linux, MacOS
- Architectures: x86_64, arm64 (MacOS-only)

Supported Linux distributions

- Ubuntu 20/22
- Debian 11/12
- Fedora 37/38
- CentOS 8
- OpenSUSE 15.4
- Mageia 8
- ArchLinux

... and FreeBSD

Thank you for your attention!
Any questions?